

# Conferenza GARR\_09

## *Selected papers*



Network Humanitatis

Linguaggi, risorse, persone

Napoli, 30 settembre - 2 ottobre 2009

# Conferenza GARR\_09

## *Selected papers*



Network Humanitatis

Linguaggi, risorse, persone

Napoli, 30 settembre - 2 ottobre 2009

ISBN 978-88-905077-0-0

Tutti i diritti sono riservati ai sensi della normativa vigente.

La riproduzione, la pubblicazione e la distribuzione, totale o parziale, di tutto il materiale originale contenuto in questa pubblicazione sono espressamente vietate in assenza di autorizzazione scritta.

Copyright © 2010 Associazione Consortium GARR

Editore: Associazione Consortium GARR

Via dei Tizii, 6, 00185 Roma, Italia

<http://www.garr.it>

Tutti i diritti riservati.

Curatori editoriali: Federica Tanlongo, Maddalena Vario, Carlo Volpe

Progetto grafico e impaginazione: Carlo Volpe

Prima stampa: Giugno 2010

Numero di copie: 1300

Stampa: Tipografia Graffietti Stampati snc

S.S. Umbro Casentinese Km 4.500, 00127 Montefiascone (VT), Italia

Tutti i materiali relativi alla Conferenza GARR 2009 sono disponibili all'indirizzo: <http://www.garr.it/conf09>.

# Indice



Introduzione .....	5
Gabriella Paolini	
Elogio dello Sperpero, del Gioco e della Pigrizia .....	7
Joy Marino	
ASTRA Project: un ponte fra Arte/Musica e Scienza/Tecnologia .....	11
Francesco De Mattia	
neuGRID: l'infrastruttura a griglia per i neuroscienziati .....	15
Alberto Redolfi	
L'accesso all'informazione scientifica online ed il sistema IDEM per l'autenticazione federata in rete .....	19
Moreno Curti	
Conservazione attiva e restauro dei supporti fonografici: il progetto REVIVAL della Fondazione Arena di Verona .....	22
Sergio Canazza	
Progetto di un sistema per la conservazione a lungo termine e a norma di legge di documenti elettronici.....	27
Raffaele Conte, Andrea Renieri	
Digital libraries su Grid per preservare e diffondere il patrimonio umanistico .....	31
Antonio Calanducci, Jorge Sevilla, Monica Saso, Alessandro De Filippo, Stefania Iannizzotto, Roberto Barbera, Giuseppe Andronico, Francesco De Mattia, Domenico Vicinanza	
Identificativi Persistenti: jNBN, un'applicazione JEE per la gestione di un'infrastruttura nazionale NBN .....	37
Emanuele Bellini, Mario Incarnato, Roberto Puccinelli, Marco Spasiano	
Tecnologie basate su XML per la fruizione avanzata dei contenuti musicali .....	43
Adriano Baratè, Luca A. Ludovico, Davide A. Mauro	
Infrastruttura telematica per la gestione informatizzata del blocco operatorio .....	48
Alessandro Pepino, Tullio Cusano	
Gestione semantica ed adattativa per la fruizione dei Contenuti Culturali nel Progetto LC3.....	55
Rocco Aversa, Beniamino Di Martino, Francesco Moscato, Salvatore Venticinque, Pasquale Cantiello	



# Introduzione

Gabriella Paolini

*Chair del Comitato di Programma della Conferenza GARR 09*



La Conferenza GARR 2009 si è tenuta dal 30 settembre al 2 ottobre 2009 a Napoli, presso l'Università degli Studi di Napoli Federico II nel Complesso Universitario di Monte Sant'Angelo con il titolo "Network Humanitatis - linguaggi, risorse, persone".

Riprendendo i temi affrontati nella Conferenza GARR 2007, l'edizione del 2009 ha mantenuto il titolo "Network Humanitatis" che simboleggia l'unione fra mondi apparentemente diversi, come le discipline umanistiche, quelle scientifiche e la telematica. Le tecnologie legate ad Internet sono usate da un numero sempre maggiore di umanisti e sono sempre più percepite come strumenti essenziali per gestire la conoscenza che deriva dalla nostra storia e dal nostro patrimonio culturale.

Il programma ha cercato di sviluppare ed arricchire gli argomenti già trattati nel 2007, mantenendo il filo conduttore della luce che viaggia nei cavi ottici della rete telematica nazionale dell'Università e della Ricerca e che illumina il pensiero e le azioni dei ricercatori, dei docenti e degli studenti italiani.

La conferenza annuale GARR riunisce gli utilizzatori della rete GARR per discutere dei principali temi e novità che riguardano tecnologie avanzate per la rete, multimedialità e innovazione. A questa edizione hanno partecipato circa 230 persone dal mondo della ricerca e accademico italiano, con una copertura nazionale dal Friuli Venezia Giulia alla Sicilia.

Quando abbiamo iniziato a pensare alla

Conferenza GARR 2009, il Comitato di Programma ha considerato che ormai la rete fosse trasversale a tutti gli ambiti di ricerca e formazione in Italia e che GARR, come NREN italiana inserita nel contesto mondiale delle reti della ricerca, giocasse un ruolo cruciale nella diffusione e nello sviluppo della rete.

GARR, come le altre NREN, fornisce un supporto al lavoro degli Enti collegati, sia come struttura di produzione, che permette il collegamento fra loro e verso l'esterno, sia come piattaforma di sviluppo e di ricerca per l'Internet del Futuro.

In quest'ottica, la ricerca sulle infrastrutture di rete, fino alle nuove applicazioni multimediali, consente alle Università e agli Enti di Ricerca italiani di essere competitivi nei contesti internazionali.

Considerando tutto questo, il Comitato di Programma ha cercato di portare nei tre giorni di conferenza le esperienze degli utenti della rete GARR di particolare interesse in campo umanistico, artistico, biomedico.

Si è parlato di calcolo distribuito per la cultura, di comunità virtuali, accesso ai dati, condivisione dei contenuti, e-learning ed Internet TV. Nella sessione di apertura hanno dato i loro spunti alla discussione Joy Marino e Gino Roncaglia.

Gli interventi inseriti nel programma sono stati in gran parte frutto della *call for papers* distribuita all'interno della Comunità GARR. Tutte le presentazioni e le registrazioni delle

riprese mandate in streaming in diretta sono disponibili sul sito web della Conferenza:  
<http://www.garr.it/conf09>.

Una selezione degli articoli presentati durante la Conferenza è presentata in questa speciale pubblicazione.

Vorrei ringraziare tutti i membri del Comitato di Programma e dell'Advisory Committee per il loro fondamentale contributo alla buona riuscita della Conferenza. È stato un piacere lavorare con loro.

Hanno fatto parte del Comitato di Programma:

- Claudio Allocchio (*GARR*),
- Giuseppe Barbieri (*Università Ca' Foscari Venezia*),
- Rosalba Belibani (*Università degli Studi di Roma La Sapienza*),
- Laura Leone (*GARR*),
- Marco Marletta (*GARR*),
- Leonardo Merola (*Università degli Studi di Napoli Federico II – INFN*),
- Laura Moltoed (*CNR*),
- Francesco Palmieri (*Università degli Studi di Napoli Federico II*),
- Joanne Maria Pini (*Conservatorio di Musica G. Verdi di Milano*),
- Mario Reale (*GARR*),
- Federica Tanlongo (*GARR*),
- Sabrina Tomassini (*GARR*).

Hanno fatto parte dell'Advisory Committee:

- Rossella Caffo (*ICCU*),
- Massimo Casciello (*Ministero della Salute*),
- Lidia Di Minco (*Ministero della Salute*),
- Massimo Parovel (*Conservatorio di Musica G. Tartini di Trieste*),
- Enzo Valente (*GARR*).

Un ringraziamento particolare va infine allo staff del Centro d'Ateneo per i Servizi Informativi (CSI) dell'Università degli Studi di Napoli Federico II che, insieme allo staff del GARR, ha curato l'organizzazione della Conferenza.



Il Complesso Universitario di Monte Sant'Angelo (sede dell'evento) e il manifesto della Conferenza GARR\_09

# Elogio dello Sperpero, del Gioco e della Pigrizia

Joy Marino

*Presidente MIX - Milan Internet Exchange*



## 1. Introduzione ed avvertenza

Ho approfittato dell'occasione che GARR mi ha offerto invitandomi a questa Conferenza per poter esprimere in piena libertà alcune idee che ho maturato negli ultimi mesi.

Sono stato sollecitato da due istantanee che ho fotografato mentalmente in due momenti assolutamente non correlati, ma che ho poi rielaborato in questo "elogio", che sembra quanto di più lontano si possa immaginare dal mondo dell'amministrazione dei soldi pubblici e della ricerca. Quest'insieme di considerazioni che, spero, mostrerà una sua coerenza nonostante l'eterogeneità dei temi, vorrei proporlo in forma di album di famiglia: una sequenza di ritratti, antichi e moderni, di persone vere.

Prima di procedere su questo terreno minato, però, corre l'obbligo di iniziare con un'avvertenza, perché è sempre facile per noi, convinti sostenitori del Genio Italiano, privilegiare questo su qualunque attività faticosa e sistematica. Noi italiani ammiriamo il modello "genio e sregolatezza", meglio se accompagnato da un pizzico di furbizia, incarnato ad esempio da Maradona, e lo preferiamo, per rimanere in ambito calcistico, alla figura del grande lavoratore, magari dai piedi meno buoni, che si dà da fare altruisticamente per tutto il campo, come il mitico "Ringhio" Gattuso. Ricordiamo invece quello che diceva Thomas A. Edison: il genio è per 1% ispirazione e per il 99% sudore (*perspiration* in inglese).

## 2. Perché un elogio

Ritornando ad Edison, uno dei punti di partenza di queste considerazioni riguarda proprio il "genio", anche nella sua accezione più prosaica di "innovazione": che cosa occorre per inventare, o anche solo per innovare, nella tecnologia, nell'industria, nella ricerca?

La prima delle due istantanee a cui accenno sopra nasce dall'incontro con un manager, anzi un manager della ricerca, il quale, di fronte all'illustrazione di un progetto complicato, articolato e sicuramente innovativo, ma ancora vago nei suoi connotati, chiedeva "Dov'è il Master Plan?"

Può essere comprensibile che chi amministra quattrini ed eroga finanziamenti si chieda come questi vengano impiegati; anch'io, in quanto amministratore di una società, seguo

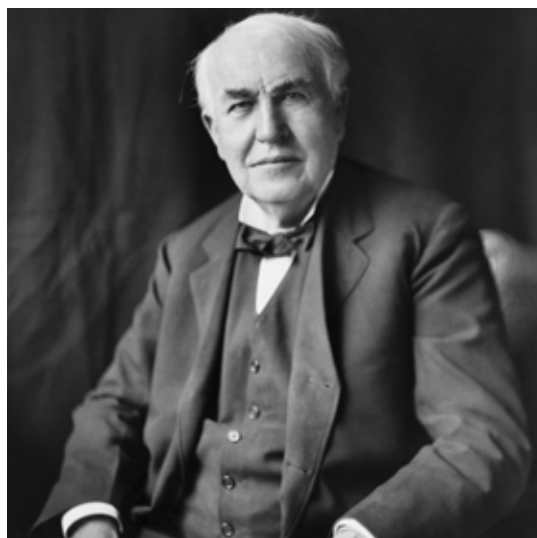


Fig. 1 Thomas A. Edison



la preparazione del Piano Strategico aziendale e del budget annuale. Come ci hanno insegnato alle scuole di formazione manageriale, se si vuole costruire una cattedrale è importante che tutti condividano gli obiettivi e si sentano parte di un piano grandioso; però, se l'obiettivo non è la costruzione di una cattedrale? Oppure: se l'obiettivo non è ancora chiaro?

### 3. Elogio dello sperpero

L'idea di Master Plan nasce all'inizio del XX secolo, e ci riporta all'inventore di quei diagrammi (Henry Gantt) che ancora oggi tutti noi utilizziamo per pianificare ed ottimizzare progetti complessi. Uno strumento utile e talvolta indispensabile, specie se l'obiettivo è l'ottimizzazione delle risorse. Il termine "Master Plan" viene anche utilizzato (vedi Wikipedia) come riferimento al "Disegno Divino", spesso dai sostenitori dell' "*Intelligent Design*" che si contrappongono all'evoluzionismo.



Fig. 2 Henry Gantt

Però forse non tutti sanno che Gantt era allievo ed amico di Frederick W. Taylor, quello del taylorismo (e fordismo) come sistema per la

produzione di massa dell'era industriale di un secolo fa: sistema di produzione che mal si adatta a questa "Società dell'Informazione".

Noi viviamo in un mondo dove l'innovazione è uno degli elementi più importanti di competizione, dove le idee, la conoscenza, il *know-how* (ma anche il design, la moda, l'effimero) sono beni essenziali per il successo di un Paese o di una azienda. In questo contesto, parlare prematuramente di Master Plan "sterilizza" la possibilità di seguire nuove strade, le quali, essendo nuove, sono necessariamente a rischio di fallire e quindi di sperperare risorse.

Questi concetti sono stati ben espressi da Clayton Christensen, che ne "The Innovator's Dilemma" mostra come una grande azienda sia incapace di produrre "innovazione dirompente", perché ormai troppo focalizzata ad ottimizzare il processo della produzione corrente.

Innovare comporta cercare strade nuove, che inizialmente saranno necessariamente in salita, vuol dire prendersi dei rischi e, spesso, fallire. Con conseguente spreco di risorse.

Ogni fallimento è uno spreco di risorse, intellettuali e materiali, ma è un processo inevitabile e necessario per il progresso. Non a caso nel mondo delle idee - così come in quello delle imprese - vige il darwinismo assoluto: la competizione, la sopravvivenza del più adatto e la ricerca di nuove opportunità o nicchie ecologiche. Da questa prospettiva, la Selezione Naturale (Charles Darwin) appare come uno splendido esempio di spreco di risorse, rispetto all'efficienza di un "*Intelligent Design*". In realtà è proprio questo il modo con cui si evolve la natura: attraverso un dispendioso processo di sperimentazione e selezione.

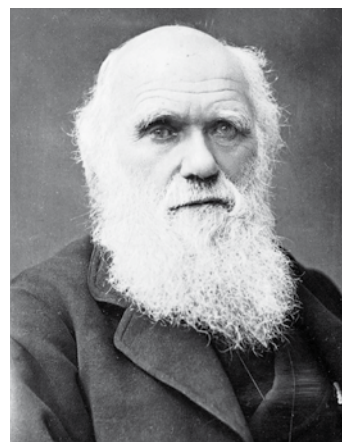


Fig. 3 Charles Darwin

### 4. Elogio del gioco

Il passaggio dallo sperpero al gioco non è ovvio (se tralasciamo qualsiasi riferimento ai *bon vivant* da *Belle Epoque*). Parlando di genio, di invenzione, di innovazione, si scopre spesso che le motivazioni che guidano la persona - appunto, "di genio" - non sono quelle dell'affermazione economica o sociale, quanto un certo atteggiamento giocoso nei confronti della vita o il desiderio di risolvere problemi di per sé; quindi, da un punto di vista

tayloristico, uno sperpero per la società intera.

Ci sono esempi innumerevoli (molti matematici ad esempio) ma, non volendo sconfinare dal campo che mi è più familiare, vorrei ricordare i due ricercatori del Bell Labs che proprio 50 anni fa concepirono e realizzarono il sistema operativo UNIX: Dennis Ritchie e Ken Thompson. Gli ingredienti alla base di quella “invenzione” furono: a) un minicomputer PDP11 poco usato, b) tanto tempo libero, c) il desiderio di avere un ambiente informatico in cui “far girare” quei giochini che potevano usare solo su un terminale remoto del Multics del MIT. Ci sono tutti gli elementi-chiave di questo mio ragionamento: sperpero (di risorse e di tempo), propensione al gioco ed un pizzico di pigrizia.

Ricordo che negli stessi anni IBM subiva il suo primo “Titanic”: dopo due anni di lavoro di centinaia di programmatori il nuovo sistema operativo per mainframe IBM360 veniva cancellato completamente senza mai arrivare in produzione. Ecco un esempio di grandioso MasterPlan!

D’altro canto, le idee che Ritchie e Thompson distillarono dentro UNIX sono ancora vive ai nostri giorni, e diretti discendenti di quel software sono ovunque, dai server web di Internet fino all’iPhone. E, dopo trent’anni, ecco che il meccanismo si ripete: uno studente finlandese, Linus Torvald, per gioco (“Just for Fun” è il titolo del libro che poi scrisse), realizza una versione di UNIX per personal computer. Nel giro di pochi anni LINUX diventa il principale concorrente dei sistemi operativi di Microsoft.



Fig. 4 Dennis Ritchie e Ken Thompson

gnificativo nella mia formazione tecnico-scientifica) circolava la battuta che UNIX fosse stato sviluppato da programmatori pigri, per la estrema sinteticità dei suoi comandi (ad esempio: “ed” anziché “edit”). Ed un fondamento di verità doveva esserci, dato che

UNIX permetteva di automatizzare facilmente tutti quei compiti ripetitivi che ogni utente di computer deve eseguire manualmente più e più volte.

La “pigrizia” del programmatore, che possiamo estendere a tutti quelli che hanno compiti ad alta componente creativa (così come ci viene richiesto sempre di più dalla Società dell’Informazione), è di tipo particolare: non è tanto un “dolce far niente”, quanto un aborreire e, se possibile, evitare tutte le attività ripetitive.

Rimanendo sempre nel campo dell’informatica, vorrei citare Larry Wall, un “*gifted programmer*”, un programmatore molto dotato, inventore di alcuni dei linguaggi più utilizzati in rete, come PERL, a sua volta capostipite di linguaggi di programmazione come PHP, Python, Ruby, ed altri. Le doti del programmatore sono per lui:

- *Laziness* - La qualità per cui ti sforzi di ridurre il consumo di energia complessivo. Che ti fa scrivere programmi salva-tempo che verranno trovati utili e documentare quello che scrivi in modo da non avere troppe domande a cui dover rispondere.
- *Impatience* - La rabbia che provi quando il calcolatore è lento. Questo ti fa scrivere programmi che non rispondono semplicemente alle tue necessità, ma le prevencono. O almeno così sembra.
- *Hubris* - L’orgoglio smodato, quel tipo di cosa per cui Giove ti può fulminare. Ma è anche la qualità che ti fa scrive-



Fig. 5 Larry Wall

## 5. Elogio della pigrizia

Molti anni fa (UNIX è stato un elemento si-

gnificativo nella mia formazione tecnico-scientifica) circolava la battuta che UNIX fosse stato sviluppato da programmatori pigri, per la estrema sinteticità dei suoi comandi (ad esempio: “ed” anziché “edit”). Ed un fondamento di verità doveva esserci, dato che UNIX permetteva di automatizzare facilmente tutti quei compiti ripetitivi che ogni utente di computer deve eseguire manualmente più e più volte.

re (e mantenere) programmi di cui non si potrà parlare male.

Questo tipo di “*laziness*”, di pigrizia, ha un valore positivo, produce idee, libera tempo per la creatività.

## **6. Conclusione**

Devo ringraziare Consortium GARR per avermi dato l’opportunità di conoscere ed apprezzare Bill Johnston, ricercatore del Lawrence Berkeley National Laboratory, che è stato Direttore di ESnet, la rete di ricerca del Dipartimento dell’Energia degli Stati Uniti e che è da poco andato in pensione.

L’estate scorsa, conversando su GARR e sulle reti della ricerca in generale, Bill ha fatto un’osservazione che condivido appieno, e questa è la seconda istantanea da cui sono partito per coagulare queste idee: “È una buona cosa che i giovani ricercatori siano esposti ad una abbondanza di risorse tecnologiche”. Si tratta di “essere esposti” (alle risorse), non di farne uso per finalità precise e predeterminate, e si tratta di “risorse in abbondanza”, probabilmente in eccesso rispetto a quanto strettamente necessario per raggiungere gli obiettivi programmati.

# ASTRA Project: un ponte fra Arte/Musica e Scienza/Tecnologia

Francesco De Mattia

*Conservatorio di Musica "A. Boito" di Parma*



**Abstract.** L'articolo illustra l'esperienza del progetto ASTRA, innovativo caso di utilizzo di risorse di Grid interconnesse dalle reti della ricerca in campo umanistico. Grazie ad un algoritmo capace di riprodurre modelli fisici degli strumenti musicali, ASTRA è in grado di restituirci il suono di strumenti andati perduti, di cui ci sono rimasti solamente frammenti, immagini e descrizioni. Gli strumenti ricostruiti possono essere suonati come qualsiasi strumento campionato, ad esempio attraverso una semplice tastiera MIDI. La ricerca si fonde quindi alla performance artistica: gli autori hanno infatti costituito uno speciale *ensemble*, Lost Sounds Orchestra, che è la prima formazione al mondo ad utilizzare strumenti antichi ricostruiti tramite ASTRA, restituendo al pubblico timbri e sonorità scomparse.

Il progetto ASTRA (*Ancient instruments Sound/Timbre Reconstruction Application*) [1] ha come scopo primario il recupero del patrimonio sonoro delle tradizioni musicali a partire dalla sintesi del modello fisico degli strumenti, dunque la riproduzione del loro timbro, mediante l'utilizzo dell'infrastruttura Grid. Si tratta, in breve, di restituire all'umanità pezzi della sua storia andata perduta, ritotendendo i suoni di quegli strumenti che per il loro contesto, oltre all'ambiente nel quale si sono conservati, non sono in condizione di essere suonati direttamente. Questi strumenti, costruiti in materiale organico come il legno, canna, osso o altri deperibili, si ritrovano normalmente in frammenti, avendo perduto parte della loro struttura originaria per diverse vicende e ragioni. Ma ASTRA si occupa anche di strumenti che possono essere suonati direttamente, con un metodo che punta anche a definire nuove variabili attraverso l'analisi della loro struttura e della loro geometria sonora: ovvero studiare le caratteristiche acustiche degli strumenti musicali da una nuova prospettiva, aggiungendo nuove forme di analisi a problemi complessi e finora poco indagati. Ciò premesso, vediamo in che modo e perché l'utilizzo della tecnologia Grid è fondamentale per la riuscita del progetto.

La tecnica di sintesi del modello fisico ha

come scopo l'uso di una serie di equazioni e algoritmi che descrivono i materiali utilizzati negli strumenti per generare una fonte fisica del suono. In altre parole, l'idea di base è di ricreare un modello dello strumento musicale e produrne il suono, simulando il suo comportamento come sistema meccanico. Ciò porta alla sintesi di un suono molto naturale, che riproduce correttamente la meccanica dello strumento. Quindi, un modello di strumento nell'ambito di ASTRA consiste in leggi della fisica che governano la produzione e l'evoluzione del suono, ed avrà vari parametri, alcuni dei quali sono delle costanti che descrivono i materiali fisici e le dimensioni dello strumento, mentre altri sono funzioni dipendenti dal tempo, descritte dall'interazione dell'esecutore con lo strumento, come colpire le corde, o coprirne i fori con le dita.

La sintesi del modello fisico non è un concetto nuovo nell'acustica, ma il suo uso pratico è diventato fattibile soltanto di recente. Si tratta infatti di una tecnica molto intensiva dal punto di vista computazionale, perché i modelli complessi degli strumenti musicali sono risolti mediante l'integrazione di equazioni differenziali. Per raggiungere la potenza computazionale necessaria e accelerare la ricostruzione del suono, pertanto, il progetto ASTRA si è spostato da un cluster di com-

puter locali a risorse computazionali distribuite. I primi test sono stati condotti con successo nell'ambito di GILDA [2] e poi su EUMEDGRID [3] mediante le reti della ricerca GÉANT [4] e EUMEDCONNECT [5]. Si può affermare che il progetto ASTRA è reso possibile solo dall'altissima affidabilità della rete GÉANT e delle reti nazionali della ricerca (NREN), dalle loro performance e dal numero di istituzioni di ricerca collegate. In questo progetto utilizziamo come input i dati archeologici (per esempio reperti provenienti da scavi, descrizioni scritte, dipinti su urne funerarie), così che la complessa tecnica di riproduzione digitale del suono venga utilizzata per modellare le caratteristiche dello strumento sotto esame. Ne consegue un modello virtuale dello strumento antico, descritto come un algoritmo che genera la forma d'onda risultante.

Dal punto di vista dell'utilizzo, l'utente di ASTRA effettua il login attraverso la *Grid User Interface*; dopo essersi autenticato, è abilitato a lanciare l'algoritmo della ricostruzione, eseguendo ASTRA attraverso i processori che costituiscono l'infrastruttura EGEE [6]. Per assistere gli utenti, che si immagina essere costituiti principalmente da musicisti e ricercatori dell'area umanistica, è stata sviluppata una interfaccia web basata sul portale GENIUS [7]. L'interfaccia di GENIUS fornisce agli utenti finali uno strato trasparente e di astrazione per accedere ai servizi e risorse distribuiti di una tipica infrastruttura a griglia.

Nella figura 1 è rappresentato il flusso di lavoro di ASTRA: l'utente effettua il download dei campioni, che possono essere caricati in un *sampler*. ASTRA al momento gira su Kontakt della Native Instruments.

Da questo momento l'utente può

suonare lo strumento ricostruito alla stregua di qualsiasi altro strumento campionato: pilotandolo ad esempio da una *MIDI keyboard* o ricevendo l'input da strumenti tradizionali ai quali vengono applicati dei *trigger*, microfoni in grado di convertire l'input audio in output MIDI.



Fig. 2 Immagine 3D dell'epigonion

L'epigonion è il primo strumento ricostruito in ASTRA. Si tratta di un antico strumento a corde simile ad un'arpa o a un salterio moderni. Fu inventato, o perlomeno introdotto in Grecia, da Epigonus di Ambracia, nell'Epiro, che fu il primo a suonare le corde con le dita, e non con un plectro.

Epigonus era un esecutore talentuoso e probabilmente portò lo strumento da Alessandria. Lo suonava con le dita di entrambe le mani, non solo come accompagnamento alla voce, cosa peraltro comune in quel periodo, ma introducendo anche passaggi cromatici e un coro di strumenti a corde,

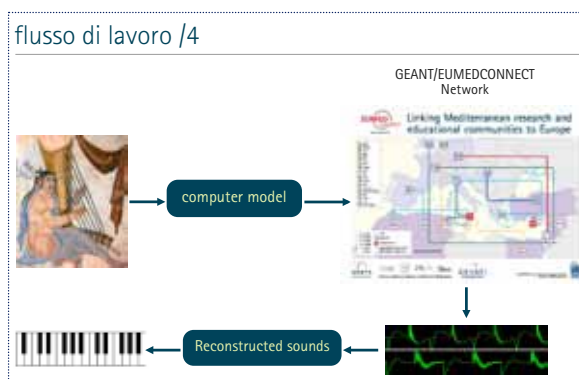


Fig. 1 Flusso di lavoro di ASTRA



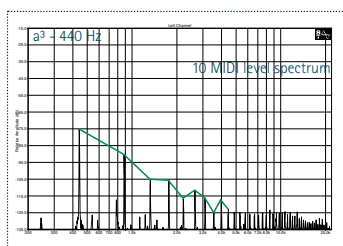


Fig. 3 10 MIDI level spectrum

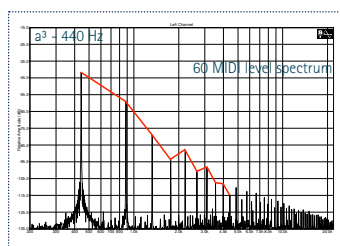


Fig. 4 60 MIDI level spectrum

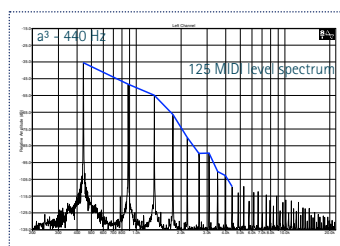


Fig. 5 125 MIDI level spectrum

probabilmente *citharas*, per accompagnarlo. Il tempo di computazione necessario per ricostruire l'epigonion è stato di circa 10 ore (reali, non teoriche) contro le 2700 stimate su computer standard; sono stati generati per ogni nota 127 diversi campionamenti, relativi ai 127 livelli di intensità del protocollo MIDI. Inoltre l'estensione è stata ampliata dalle originali 40 a 56 note: 7112 campioni, per un totale di circa 20 GBytes.

Le figure 3, 4 e 5 raffigurano gli spettri di una nota  $La_3$  a 440 Hz suonata con tre diverse intensità crescenti. Nella figura 6 sono stati sovrapposti i tre spettri: si nota come il timbro cambia a seconda dell'intensità del tocco, come nello strumento reale, per cui l'esecutore può cambiare la dinamica rendendo l'esecuzione molto naturale. Grazie a questa caratteristica, abbiamo verificato che gli strumenti ricostruiti in ASTRA si fondono perfettamente con gli strumenti reali, soprattutto dell'area barocca, anche se non disdegniamo la sperimentazione di nuove sonorità. A questo proposito abbiamo costituito uno speciale *ensemble*, Lost Sounds Orchestra [8], che è l'orchestra ufficiale di ASTRA Project. È la prima orchestra al mondo ad utilizzare strumenti antichi ricostruiti tramite la sintesi del modello fisico.

L'esperienza di ASTRA viene trasferita a tutte le istituzioni che partecipano ad ACHALAI,

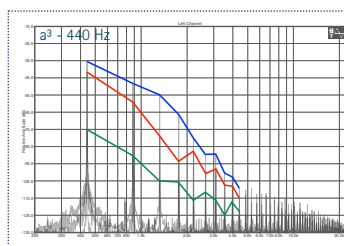


Fig. 6 I tre spettri sovrapposti

(progetto cileno che in lingua quechua significa “*che bello*”) che vede coinvolte cinque Università cilene oltre al Museo Chileno de Arte Precolombino, col supporto di REUNA, la rete accademica nazionale cilena.

Questo progetto permette il recupero degli antichi strumenti musicali preispanici della regione andina. In questo modo musicisti ed esecutori di tutto il mondo potranno riscoprire questo singolare tesoro artistico, impiegando nuovamente timbri e strumenti spariti da ormai molto tempo. L'enorme lavoro editoriale sarà curato dalla sezione editing di ASTRA, e permetterà di avere in notazione moderna convenzionale tutti i brani che saranno recuperati tramite la ricerca in archeologia musicale e archeomusicologia. Infine, è in fase di costituzione un gruppo di esperti per estendere i risultati del progetto anche agli altri paesi dell'America Latina. Possiamo affermare che ASTRA ha dato un formidabile impulso alla costituzione in quel continente di un gruppo di ricerca dedicato alle arti che utilizza la



Fig. 7 L'epigonion suonato durante un concerto della Lost Sounds Orchestra

rete accademica quale vettore di scambio di informazioni.

L'utilizzo di ASTRA in performances musicali in real-time implica il superamento di problematiche legate a latenze, *routing* dei flussi audio e digitale, riverberi, e comunque al setup in generale. Per questo, dall'inizio di settembre 2009 ASTRA Project ha ottenuto il sostegno logistico e il supporto tecnico di Gribby Music, importatore per l'Italia di prestigiosi marchi per l'audio professionale.

Concludendo, in progetti di ricerca come ASTRA la rete ricopre davvero un ruolo fondamentale, quello di aprire spazi per lo sviluppo delle idee, attraverso lo sforzo congiunto, nella generazione e nella trasmissione della conoscenza. La capacità di integrazione tra i gruppi di ricerca a livello mondiale, diven-

ta così anche garanzia di qualità ed efficienza della produzione scientifica locale.

### Riferimenti bibliografici

- [1] <http://www.astraproject.org>
- [2] <https://gilda.ct.infn.it/>
- [3] <http://www.eumedgrid.org>
- [4] <http://www.geant.net>
- [5] <http://www.eumedconnect.net>
- [6] <http://www.eu-egee.org/>
- [7] <https://grid.ct.infn.it/twiki/bin/view/GILDA/WebHome>
- [8] <http://www.lostsoundsorchestra.org>

Le immagini 3D dell'epigonion sono state realizzate da Francesca Baghino e Fausto Ugozoli @ VisArc Studio, Parma.

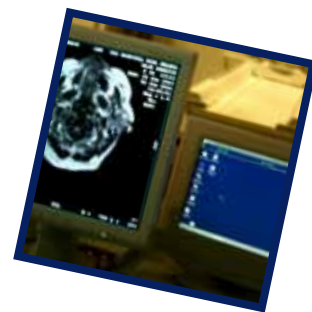


**Fig. 8** The Lost Sounds Orchestra in concerto

# neuGRID: l'infrastruttura a griglia per i neuroscienziati

Alberto Redolfi

*IRCCS Centro S. Giovanni di Dio - Fatebenefratelli di Brescia  
Consorzio neuGRID*



**Abstract.** L'articolo presenta le applicazioni del progetto neuGRID nello studio delle malattie neurodegenerative, come la malattia di Alzheimer, grazie all'utilizzo dell'infrastruttura di Grid.

Oggi infatti le comunità di neuroimaging si avvalgono sempre più di strumenti statistici e matematici per estrarre informazioni da immagini di tessuti biologici ed in poco tempo la crescente disponibilità di immagini e dati comporterà la necessità di aumentare le risorse.

Il progetto neuGRID si propone di raccogliere e archiviare grandi quantità di dati di imaging e di analizzarle utilizzando un'infrastruttura di Grid distribuita capace di una enorme potenza di calcolo e di uno spazio di accumulo dati pressoché illimitato.

L'infrastruttura di Grid interconetterà gli ospedali di tutta Europa e offrirà ai neuroscienziati e ai ricercatori le più avanzate tecnologie di comunicazione per superare la patologia di Alzheimer.

neuGRID è un progetto triennale interamente finanziato dalla Comunità Europea nell'ambito del 7° Programma Quadro. Il progetto prevede lo sviluppo di un'infrastruttura digitale, basata sul sistema Grid, che permetterà alla comunità di neuroscienziati di poter avanzare nella ricerca scientifica per lo studio delle malattie neurodegenerative e, in particolar modo, nella malattia di Alzheimer.

neuGRID è un progetto che fonde al proprio interno differenti campi di ricerca conferendo all'infrastruttura una unicità indiscussa nel campo delle neuroscienze. neuGRID, infatti, nasce dall'integrazione di competenze nell'ambito delle innovative tecnologie di imaging strutturale, conoscenze in ambito di ricerca clinica e nell'ambito delle neuroscienze, nonché un massiccio impiego di infrastrutture computazionali ad alte prestazioni.

Oggigiorno l'esigenza di una potente infrastruttura a griglia in grado di fornire in breve tempo risultati quantitativi a neuroscienziati e ricercatori è di primaria importanza. Infatti, nella malattia di Alzheimer, siamo ormai molto vicini a definire marcatori significativi per il monitoraggio di questa forma di demenza e, in base alla definizione di marca-

tori affidabili, potremo determinare l'efficacia di farmaci in grado di modificare il decorso della malattia. In aggiunta, i nuovi marcatori che verranno caratterizzati potranno poi supportare i neurologi anche in fase di diagnosi della malattia, consentiranno di misurarne l'evoluzione, permetteranno di valutare l'efficacia di farmaci e trattamenti, consentiranno di migliorare ulteriormente la comprensione dei meccanismi di malattia e le funzioni cerebrali compromesse. Un esempio innovativo di marcatore per la malattia di Alzheimer è lo spessore della corteccia cerebrale. Nel tempo, le variazioni di questa struttura possono fornire informazioni importanti per i neurologi. In tal senso, la piattaforma neuGRID mette a disposizione dei ricercatori uno degli algoritmi più consistenti ed efficaci per la quantificazione dello spessore corticale. neuGRID, inoltre, fornisce una quantità di dati che non è mai stata disponibile in passato per effettuare studi significativi di efficacia. Oggi, invece, con più di 8000 scansioni di risonanza magnetica sarà possibile migliorare e raffinare maggiormente la comprensione di questi innovativi marcatori di malattia. neuGRID con il proprio portafoglio di algoritmi (basti citare FSL, MRICron, Free-



Surfer, MNI-BIC-CLASP) e *pipeline* (insiemi di elaborazioni eseguite da differenti software e/o algoritmi) renderà possibile tutto ciò che fino ad una manciata di anni fa era impensabile. La necessità dell'impiego di griglie all'interno del progetto neuGRID è giustificata dal fatto che le *pipeline* di analisi messe a punto dagli sviluppatori di algoritmi, dai bioinformatici o anche dagli stessi ricercatori risultano essere molto onerose sia da un punto di vista computazionale che di gestione dei dati. Tipicamente, infatti, in questi tipi di analisi vengono impiegate migliaia di immagini di risonanza magnetica cerebrale caratterizzata ognuna da una dimensione di 20-40 megabyte.

Appare evidente che l'obiettivo dell'infrastruttura neuGRID è quello di cambiare radicalmente il paradigma con cui tipicamente i ricercatori, nel campo delle neuroscienze, opereranno nella ricerca di tutti i giorni. Ad oggi, per riuscire ad estrarre evidenze significative sui propri *dataset* di immagini, lo scenario tipico è che un ricercatore faccia una copia digitale dei propri dati e si rechi nei più grossi centri di analisi di neuroimaging (ad esempio il "LONI" presso l'Università della California a Los Angeles o presso il Montreal Neurological Institute - MNI, in Canada). In questi centri, dotati dei più innovativi algoritmi di analisi e di specifiche infrastrutture, il ricercatore solitamente può imparare la metodica, per poi esportarla nel proprio centro di ricerca. A questo punto il ricercatore dovrà essere in grado, quando possibile, di installare e allestire le procedure di analisi apprese all'estero anche presso i propri cluster di analisi. Come è evidente, lo scenario sopradescritto può mostrare grosse limitazioni (ad esempio, incompatibilità di differenti infrastrutture, ottimizzazione dei processi d'analisi, ecc.). neuGRID, per contro, propone un approccio più lineare ed esattamente opposto a quello fino ad oggi utilizzato. neuGRID infatti porta il laboratorio, le infrastrutture di calcolo, gli algoritmi e le *pipeline* di analisi direttamente a casa del ri-

cercatore. Ciò è reso possibile sfruttando un efficace portale web in grado di interfacciare facilmente l'utente finale all'infrastruttura di analisi neuGRID.

Da un punto di vista tecnico neuGRID propone una *Service Oriented Architecture* (SOA) basata su middleware EGEE/gLite. L'architettura è caratterizzata da tre principali classi di servizi:

- *Servizi per l'utente finale.* Questi sono servizi specifici e solo parzialmente riutilizzabili in altri progetti simili a neuGRID (ad esempio, Visualization Service e Image Analysis Service).
- *Servizi medici generici.* Questi servizi potrebbero essere impiegati in altre applicazioni di carattere medico (ad esempio, Medical Querying Service e Provenance Service).
- *Servizi specifici per la griglia.* Sono servizi generici che potrebbero essere completamente riutilizzabili in qualsiasi altro progetto in grado di utilizzare le infrastrutture a griglia (ad esempio, Computing Service e Storage Service).

Come appare chiaro, la filosofia che sottende neuGRID è quella che abbraccia i principi di flessibilità, espandibilità e di riutilizzabilità: caratteristiche chiave per garantire il successo della piattaforma neuGRID anche in futuro.



Fig. 1 Infrastruttura neuGRID

In aggiunta, neuGRID è caratterizzato da un'infrastruttura gerarchica composta da due livelli principali (livello 0 e livello 1). Il livello 0 è il livello più alto della griglia. Que-

sto ospita sia il centro di coordinamento della griglia che quello che coordina e gestisce i dati custoditi in neuGRID. Il livello 1 è caratterizzato invece dai centri di calcolo veri e propri (IRCCS Fatebenefratelli, Karolinska Institute, VU Medical Centre). In pratica, tutti gli utenti finali possono accedere alla potenza di calcolo del livello 1 utilizzando il portale neuGRID (<http://neugrid.healthgrid.org>) e sfruttando la suite di applicativi messi a loro disposizione (fig. 1).

Ad oggi, l'infrastruttura neuGRID è attiva al 100%. In particolare, sia il livello 0 che il livello 1 sono completamente configurati e funzionanti e, conseguentemente, l'intera potenza di calcolo è usufruibile da parte dell'utente finale. Entro la fine del 2010 saranno conclusi tutti i lavori di ottimizzazione dell'interfaccia e, successivamente, la griglia neuGRID sarà pronta per essere testata e ulteriormente raffinata entro la fine del 2010. Infine, da un punto di vista infrastrutturale neuGRID è connesso alla rete GARR e alla rete paneuropea GÉANT.

neuGRID consentirà di effettuare analisi in maniera semplice, efficiente ed efficace. Infatti, una volta che i neuroscienziati e neuroriceratori ottengono il proprio certificato, possono accedere tramite il portale web di neuGRID ai servizi resi disponibili. A tal proposito, in neuGRID è stata adottata la tecnologia Web 2.0 Ajax-based e il sistema di *single sign-on*, in modo tale da consentire all'utilizzatore finale di eseguire complesse analisi sfruttando la rapidità e la potenza di calcolo della griglia. Tutto il sistema neuGRID è stato pensato integrando le più rigorose politiche di sicurezza e gli standard web più innovativi oggi utilizzabili.

Una volta acceduto al sistema, e prima di far correre e disegnare le proprie *pipeline* di analisi di immagini di pazienti affetti da patologie cerebrali, il neuroscienziato può definire e selezionare la propria popolazione di studio e/o di riferimento. In neuGRID tutto ciò è reso possibile sfruttando un efficiente sistema di database distribuito chiamato LORIS. Questo



**Fig. 2** Pannello superiore: Portale neuGRID;  
Pannello inferiore sinistro: Loris Dataset;  
Pannello inferiore destro: Immagine anonimizzata.

sistema fornisce, infatti, la possibilità di effettuare controlli di qualità preventivi sull'immagine che si andrà ad analizzare e consente di formulare delle specifiche query per poter interagire con l'intero archivio costituito da più di 8000 scansioni cerebrali.

Inoltre, trattando neuGRID dati sensibili, si è cercato di porre molta attenzione anche all'impiego e alla gestione di tali dati e delle relative informazioni ad essi associati. È importante sottolineare che qualsiasi acquisizione cerebrale, prima di essere archiviata in LORIS, subisce una opportuna anonimizzazione e una deturpazione delle caratteristiche biometriche del volto al fine di garantire la privacy di ogni soggetto archiviato nel sistema neuGRID (fig. 2).

In seguito, l'utente, una volta autenticatosi e scelto il campione di studio, potrà essere interessato a sottomettere le proprie analisi ai *Computing Elements* (CE) della griglia. neuGRID, per venire incontro alle esigenze di differenti utenti finali, offre due diverse possibilità per effettuare il processamento delle immagini: utilizzare l'interfaccia a riga di comando tramite un'opportuna *shell* (GSI-SSH), oppure, utilizzare una *Graphical User Interface* (GUI) evitando la scrittura di qualsiasi riga di codice. Ciò risulta essere molto utile so-

prattutto per utenti meno esperti. Nel primo caso, tramite il web portal, è possibile invocare la GSI-SSH per poter accedere in maniera diretta alle *pipeline* di analisi già gridificate e ad algoritmi opportunamente installati. Oppure, considerato che i comandi gLite dell'ambiente Grid possono essere ostici da apprendere da parte dell'utente finale, in neuGRID è stata sviluppata un'apposita GUI che consente di sottomettere *pipeline* d'analisi gridificate e algoritmi semplicemente con un click del mouse.

Recentemente, come prova di funzionamento e di verifica dell'infrastruttura neuGRID, è stata gridificata la *pipeline* CIVET, che consente di valutare l'atrofia del mantello corticale attorno al cervello (fig.3). Come ricordato precedentemente, questo è considerato oggi un marcatore innovativo per la malattia di Alzheimer. CIVET risultava essere molto adatto come test poiché l'analisi di una singola scansione di risonanza magnetica impiega 7 ore per essere completata usando un processore standard. CIVET rappresenta infatti una *pipeline* molto intensa per la mole di dati generati facendo registrare un'esplosione di un fattore pari a 10x se si considera l'output finale. In aggiunta, per essere opportunamente eseguito e gridificato, CIVET necessita della risoluzione di numerosissime dipendenze, pertanto tutte queste caratteristiche ne fanno un candidato ideale per creare un buon *data challenge*.

Al fine di rendere questo test di funzionalità dell'infrastruttura neuGRID ancor più significativo, sono state analizzate 6000 scansioni di risonanza magnetica nucleare ad alta intensità raccolte analizzando, per più di 3 anni, 800 soggetti (Soggetti normali, MCI e con Alzheimer). È interessante sottolineare che l'esecuzione di questo "*job*" in griglia è durato 2 settimane anziché 5 anni, tempo che sarebbe stato necessario se la medesima analisi fosse stata lanciata su un computer "*mono-core*". Infine, a seguito di questa intensiva analisi, in output è stato generato un volume di dati pari a 1,5 Terabyte, che è attualmente in fase di studio.

Grazie a questa dimostrazione di forza, l'infrastruttura neuGRID è risultata vincitrice del premio per il miglior evento dimostrativo durante la conferenza EGEE '09, tenutasi lo scorso settembre a Barcellona.

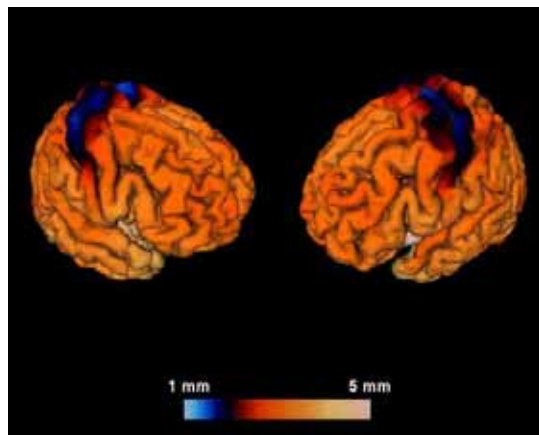


Fig. 3 Spessore corticale derivato tramite CIVET in neuGRID

In conclusione, al fine di offrire ai neuroscienziati di tutto il mondo un'infrastruttura ottimale per migliorare lo studio di malattie ad alto impatto sociale come quella di Alzheimer, neuGRID sta cercando di rafforzare maggiormente le proprie capacità di calcolo. Nuove collaborazioni sono state varate con infrastrutture di analisi analoghe localizzate in Canada, come la piattaforma CBRAIN, o negli Stati Uniti, come la nota infrastruttura LONI.

La volontà di neuGRID è quella di costituire nei prossimi anni un'infrastruttura internazionale caratterizzata da più di 6000 *core* e più di 200 Terabyte di spazio, offrendo ai ricercatori e neuroscienziati la più grande, esaustiva e completa infrastruttura di analisi e di archivio dati disponibile. Inoltre, con neuGRID si potrà ridurre notevolmente la latenza nell'ottenimento dei risultati, incrementare la produttività dei ricercatori e velocizzare il raggiungimento di nuove conoscenze. Una volta concluso, neuGRID offrirà un portafoglio consistente di servizi utilizzabili da una vasta comunità rappresentata non solo da ricercatori e neurologi, ma anche da sviluppatori di algoritmi nel campo del *brain imaging* e, verosimilmente, anche dalle industrie farmaceutiche.

# L'accesso all'informazione scientifica online ed il sistema IDEM per l'autenticazione federata in rete

Moreno Curti

Fondazione IRCCS Policlinico San Matteo di Pavia  
Coordinatore Bibliosan



**Abstract.** In questo articolo viene affrontato l'argomento dell'accesso all'informazione scientifica online tramite un'autenticazione ed autorizzazione basata su Shibboleth un sistema software open source. Attraverso una singola autenticazione (web single sign-on), il sistema permette all'utente di accedere ai servizi protetti situati sia all'interno della propria organizzazione sia in altre esterne che aderiscono ad una federazione con regole condivise. Per fare ciò, occorre realizzare una Infrastruttura di Autenticazione e Autorizzazione (AAI) che consenta di trasferire ad ogni singola organizzazione il compito di dare garanzie sulle identità e i diritti dei propri utenti. In particolare viene presentata l'esperienza di Bibliosan, la rete delle biblioteche degli enti di ricerca biomedici italiani, che ha aderito ad IDEM, l'AAI federata della rete GARR.

Uno dei problemi che sempre più frequentemente tormenta chi utilizza servizi informativi a pagamento su internet (banche dati, riviste elettroniche, etc.) è la difficoltà a gestire e ricordare un numero crescente di username e password. Non sempre la possibilità di essere riconosciuti in base all'indirizzo IP (Internet Protocol) del proprio computer permette di risolvere il problema legato all'autenticazione. Nel mondo delle biblioteche e dei centri di documentazione scientifica, per esempio, ciò è evidente quando si accede ai portali degli editori (ad esempio, Elsevier, Thomson Reuters, British Medical Journal, Sage, Jama, etc.) o quando si utilizzano servizi come NILDE (*Network Inter-Library Document Exchange*), NILDE-Utente<sup>1</sup>, Refworks<sup>2</sup> o CLAS (*Cilea Library Access Services*)<sup>3</sup>: tutti servizi ai quali sono associati differenti username e password che, per necessità di sicurezza legate ai contratti sottoscritti con i singoli fornitori di servizi, non possono essere ridotte ad una sola. Il problema può ovviamente ingigantirsi all'aumentare dei servizi e delle risorse a cui l'utente può accedere tramite autenticazione, creando non poche difficoltà per una gestione ed un utilizzo efficace ed efficiente delle proprie credenziali.

In Internet si sta diffondendo la possibilità di accedere all'informazione attraverso un'autenticazione ed autorizzazione unica (*web single sign-on*) basata su Shibboleth, un sistema software open source che permette all'utente di accedere a servizi in rete protetti, situati sia all'interno della propria organizzazione che in altre esterne. L'utilizzo di Shibboleth comporta l'adesione ad una federazione in cui sono applicate regole condivise per l'identificazione degli utenti e l'autorizzazione per l'accesso a risorse e servizi.

Nel sistema Shibboleth, accanto agli utilizzatori (*Identity Provider*) si pongono i fornitori di servizi (*Service Provider*) che, nel mondo dell'informazione scientifica, si stanno sempre più adeguando a questa nuova opportunità. Editori internazionali (quali Elsevier, Thomson Reuters, WoltersKluwer, British Medical Journal), aggregatori di risorse.....

<sup>1</sup> Servizio in rete che permette agli utenti di un ente registrato di inviare alla propria biblioteca ogni richiesta necessaria per il recupero di documentazione scientifica (*document delivery*).

<sup>2</sup> Software web-based per la gestione bibliografica personale di documentazione scientifica.

<sup>3</sup> Interfaccia web-based che consente di accedere da remoto alle risorse documentali in rete non altrimenti raggiungibili dall'esterno della propria organizzazione in quanto il metodo di autenticazione è il solo riconoscimento IP.

se (come Ebsco, Proquest), altri sistemi come DSpace (open archives), Refworks, Google Apps, NILDE, etc., hanno, infatti, iniziato a sviluppare ed offrire i propri servizi attraverso un accesso autenticato via Shibboleth.

In sostanza, Shibboleth cambia le modalità di controllo degli accessi e trasferisce alle strutture di appartenenza degli utenti il compito di dare garanzie sulla loro identità (autenticazione) e sui loro diritti (autorizzazione), mediante la realizzazione di una Infrastruttura di Autenticazione ed Autorizzazione (AAI).

In quest'ottica, in Italia, è stato realizzato il progetto IDEM (IDentity Management federato per l'accesso ai servizi), il progetto pilota proposto dal GARR ed ora diventato sistema, per la realizzazione dell'Infrastruttura di Autenticazione e Autorizzazione (AAI) federata. Questa AAI gestisce i processi di autenticazione e autorizzazione fra un utente, la sua organizzazione d'origine e la risorsa cui vuole accedere ed utilizza Shibboleth quale framework per instaurare relazioni basate sulla fiducia all'interno della comunità dei partecipanti.

Il Consortium GARR ha sottoscritto nel 2005 una convenzione con il Ministero della Salute per collegare sulla propria rete tutti gli Istituti di Ricovero e Cura a Carattere Scientifico (IRCCS). Bibliosans, in quanto rete delle Biblioteche di tutti gli Enti di Ricerca Biomedici Italiani che fanno capo al Ministero della Salute e che comprende ad oggi 56 Enti (43 IRCCS, 10 Istituti Zooprofilattici Sperimentali, l'Istituto Superiore di Sanità, l'Istituto Superiore per la Prevenzione e la Sicurezza sul Lavoro - ISPESL e l'Agenzia Nazionale per i Servizi Sanitari Regionali - ANSSR), è stata coinvolta nel 2008 nel progetto IDEM per conoscere le caratteristiche e riflettere sui potenziali vantaggi derivabili dall'utilizzo dei servizi disponibili sulla rete GARR.

Attraverso una propria struttura organizzativa, Bibliosans coordina le acquisizioni di risorse informative online ed offre alle bibliote-

che del sistema una serie di risorse e di servizi informativi a carattere scientifico che comprendono, tra l'altro, l'accesso ad oltre 5.000 riviste a pagamento disponibili via web, un software web-based per la gestione in rete della documentazione scientifica personale (RefWorks) e, sempre su web, un sistema di comunicazione tra gli utenti e le proprie biblioteche orientato a semplificare ed accelerare il sistema di recupero di documentazione quando non direttamente accessibile in rete (sistema NILDE e NILDE-Utente). Grazie a quest'ultimo software web-based, l'utente può registrarsi online presso la propria biblioteca e richiedere ad essa, dopo essersi ogni volta autenticato, quanto necessario per soddisfare il proprio bisogno documentale. I bibliotecari, a loro volta, grazie a NILDE, un software anch'esso web-based e a loro esclusivo uso, si interfacciano con le altre biblioteche del sistema e, dopo aver effettuato l'autenticazione, richiedono ad esse la documentazione necessaria ai propri utenti.

Attualmente le statistiche relative ai servizi offerti da Bibliosans hanno raggiunto numeri molto significativi, indicativi di quanto sia importante facilitare l'accesso all'informazione scientifica dall'interno delle strutture di ricerca: oltre 1,5 milioni di articoli scientifici scaricati annualmente dai siti dei principali editori biomedici, oltre 2.000 NILDE-Utenti e CLAS-utenti, oltre 1.300 utenti di Refworks, oltre 50.000 articoli all'anno recuperati con il *document delivery*.

In questo complesso sistema di servizi per la documentazione scientifica, soprattutto ora che gli utenti registrati sono diverse migliaia e stanno ulteriormente aumentando, risulta sempre più evidente il problema legato alla gestione degli account e degli accessi. È diventato oneroso persino per gli utenti stessi gestire, memorizzare, usare e riutilizzare al momento giusto più password che, considerata l'esplosione di Internet e dei servizi in essa accessibili (banche, telefonia, servizi a pa-



gamento vari, etc.), stanno aumentando e sono di volta in volta richieste. Anche all'interno delle proprie strutture di ricerca, molto spesso, si presenta lo stesso problema, in particolare quando è necessario autenticarsi per accedere a dati riservati e protetti di varia natura (dati sanitari, amministrativi, banche date, registri, etc.).

La realizzazione di un sistema Bibliosan basato su Shibboleth, che coinvolge tutte le strutture ad esso aderenti, permetterebbe di risolvere tutti i problemi di autenticazione soprariporati. Ciò richiede innanzitutto la creazione di un *Identity Provider* presso ciascuna struttura e l'adesione di tutti gli

Istituti alla Federazione IDEM del GARR.

L'adesione a questo progetto e le prime sperimentazioni sono state possibili alla Fondazione IRCCS Policlinico San Matteo di Pavia dove, per fatti contingenti, si è verificata una situazione favorevole: la creazione (indispensabile per l'implementazione locale di Shibboleth) di un sistema centralizzato di *single sign-on*, che permette ad un utente di autenticarsi una sola volta e di accedere a tutte le risorse informatiche alle quali è abilitato semplificando anche la definizione e la gestione delle password e delle politiche di sicurezza.

Grazie alla collaborazione con CILEA (Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica) e dopo aver analizzato e studiato il problema e le potenziali ricadute positive derivanti dall'implementazione di Shibboleth, è stato installato presso il San Matteo un server di *Identity Provider*

(*Software IdP*: Shibboleth versione 1.3 su Sistema Operativo Debian Etch; *Frontend*: Apache2/Tomcat 5.5; *Backend*: sistema/servizio Active Directory) e sono stati condotti i primi test mediante collegamento con un Service Provider *Shibboleth-compliant* situato presso il CILEA stesso (CLAS - Cilea Library Access Service).

Dopo i risultati positivi ottenuti ed in segui-

to alla trasformazione di IDEM da progetto a sistema, il Policlinico San Matteo di Pavia ha attivato tutte le pratiche necessarie per aderire ufficialmente alla Federazione fornendo così a tutte le strutture Bibliosan interessate, un modello funzionante e di



Fig. 1 Screenshot dal sito di Bibliosan

riferimento (fig.1) per una loro futura implementazione.

# Conservazione attiva e restauro dei supporti fonografici: il progetto REVIVAL della Fondazione Arena di Verona

Sergio Canazza

*Dipartimento di Ingegneria dell'Informazione,  
Università degli Studi di Padova*



**Abstract.** L'articolo illustra in dettaglio il progetto di ricerca REVIVAL focalizzato sulla conservazione e sul restauro dell'Archivio Audio Vicentini (Fondazione Arena, Verona).

L'apertura di archivi e biblioteche ad un'ampia comunità telematica, resa disponibile dall'integrazione nella rete Internet, ha rappresentato un impulso fondamentale allo sviluppo culturale e didattico. Le difficili condizioni di agibilità di molti archivi di documenti sonori e la loro dispersione sul territorio nazionale ne limitano pesantemente la consultazione rendendo onerosa la ricerca dei materiali. La realizzazione di un progetto di digitalizzazione dei documenti audio può oggi consentire l'accesso in rete di questo patrimonio documentale, secondo una scelta di ingegneria culturale in linea con le grandi biblioteche europee e nordamericane.

Nel decreto legislativo n. 112 del 1998 (in attuazione della legge n. 59 del 1997), al Titolo IV (Servizi alla persona e alla comunità), Capo V (Beni e attività culturali), per la prima volta viene data una precisa definizione di "tutela dei beni culturali" (art. 148 "Definizioni", comma 1, lettera c): «ogni attività diretta a riconoscere, conservare e proteggere i beni culturali e ambientali». In questo senso, nella tutela possono essere compresi, tra gli altri, i sistemi di protezione antincendio e antifurto nonché, com'è esplicitamente scritto, le operazioni di conservazione. Quest'ultime consistono nel complesso delle azioni, dirette e indirette, volte a rallentare gli effetti della degradazione causata dal tempo e dall'uso sulle componenti materiali dei beni culturali.

Nel campo dei dischi fonografici, la conservazione si articola in passiva<sup>1</sup> (difesa del supporto dagli agenti ambientali, senza alterarne la struttura) e attiva (trasferimento dei dati nel dominio digitale, spesso nota, nella letteratura anglosassone come *digital curation*). Poiché il disco è caratterizzato da un'aspettativa di vita relativamente bassa (se confrontata con quella di altri monumenti) la conservazione passiva risulta insufficiente. Inoltre, i dischi

non hanno la possibilità di vicariare o rigenerare l'informazione perduta partendo dal segnale audio: è quindi di fondamentale importanza trasferire nel dominio digitale (oltre al segnale audio) tutta l'informazione contestuale presente nel supporto. In questo senso, molti archivi inseriscono nella copia conservativa le riprese fotografiche (etichetta, supporto, allegati, corrotte) e video effettuate durante il trasferimento dall'analogico al digitale (A/D) del segnale audio (disco in rotazione). Questi documenti, infatti, offrono informazioni sulle alterazioni (intenzionali o meno) e sulle corrotte del supporto. Questi dati sono necessari a fini archivistici, per studi musicologici e nelle operazioni di restauro audio [1], [2], [3]. Nel caso di archivi di medie/grandi dimensio-

.....

<sup>1</sup> La conservazione passiva si suddivide a sua volta in indiretta, che non comporta il coinvolgimento fisico del disco, e diretta, nella quale il disco viene trattato, senza comunque alterarne struttura e composizione. Nella conservazione passiva indiretta rientrano: la prevenzione ambientale (che si esplica attraverso il controllo dei parametri ambientali che sono, in ordine decrescente di pericolosità per i dischi: umidità relativa, temperatura, inquinamento, luce), la formazione del personale addetto alla conservazione, l'educazione dell'utente. La conservazione passiva diretta comprende gli interventi di: realizzazione di custodie di protezione, spolveratura delle raccolte, disinfestazione degli archivi con gas inerti.

ni, è economicamente improponibile estrarre manualmente le informazioni di interesse dai filmati. Diventa quindi prezioso poter contare su strumenti in grado di individuare automaticamente eventi di interesse presenti nelle riprese fotografiche e nelle registrazioni video dei dischi fonografici.

L'apertura di archivi e biblioteche ad un'ampia comunità telematica, resa disponibile dall'integrazione nella rete Internet, rappresenta un impulso fondamentale allo sviluppo culturale e didattico. Le difficili condizioni di agibilità di molti archivi di musica contemporanea e la loro dispersione sul territorio nazionale ne limitano pesantemente la consultazione, rendendo onerosa la ricerca dei materiali. La realizzazione di un progetto di digitalizzazione dei documenti audio può oggi consentire l'accesso in rete di questo patrimonio documentale, secondo una scelta di ingegneria culturale in linea con le grandi biblioteche europee e nordamericane. In questa sede si intende presentare il progetto di ricerca "REstauro dell'archivio Vicentini di Verona e sua accessibilità come Audio e-Library" (REVIVAL), focalizzato sulla conservazione e sul restauro dell'Archivio Audio Vicentini (Fondazione Arena, Verona). Questo archivio possiede la potenzialità per attirare l'interesse delle comunità scientifica e musicologica internazionali, grazie alle decine di migliaia di documenti sonori in suo possesso, alcuni di grande rarità o addirittura unici al mondo. A fronte di questo grande potenziale, una recente analisi ha evidenziato che una grande parte di documenti audio sono a rischio di scomparsa: questo provocherebbe un grave danno al patrimonio culturale esistente in Italia.

REVIVAL intende condurre azioni di conservazione (attiva e passiva) dei documenti sonori e di restauro del segnale audio nei documenti di maggior interesse commerciale e realizzerà una base di dati al fine di garantire l'accessibilità ai documenti restaurati e rendere a lungo termine il sistema auto-sostenibile.

Oltre all'evidente impatto musicologico, il progetto sta sviluppando sinergie con istituzioni (pubbliche e private) locali, nazionali e internazionali, interessate a implementare su larga scala il programma di conservazione applicato all'Archivio Vicentini. Con la collaborazione della Fondazione Arena di Verona e dell'Università di Verona e con il supporto scientifico di Eye-Tech srl (Udine), il progetto si propone di realizzare una piattaforma HW/SW finalizzata alla conservazione attiva, al restauro e all'accesso dei documenti sonori dell'Archivio Vicentini di Verona.

### 1. Conservazione attiva

Sistemi di *Automatic text scanning* e *optical character recognition* vengono largamente usati nelle biblioteche; diversamente da quanto accade per i testi, però, la conversione A/D del segnale audio inciso nei dischi fonografici viene solitamente eseguita per mezzo di un processo invasivo. Esistono, tuttavia, *laser turntable* in cui il tradizionale *pickup* è sostituito da un raggio laser. In questo modo, il supporto non viene fisicamente toccato durante la lettura. Questi apparecchi soffrono però di alcuni svantaggi:

- sono molto sensibili ai graffi nel supporto e alle eventuali ondulazioni anomale del solco;
- la capacità di riflessione della superficie del disco deve essere ottima;
- sono vincolati all'informazione presente in un ridotto *laser spot*.

L'autore ha applicato alcuni strumenti sviluppati nel campo della visione artificiale per sintetizzare il segnale audio direttamente dall'immagine fonografica della superficie del disco [4], [5], [6], [7]. Questo approccio permette:

- la conservazione di tutta l'informazione presente nel supporto;
- una lettura completamente non invasiva;
- la conservazione attiva di supporti le cui corrotture ne renderebbero impossibile una lettura tradizionale;



- l'avvio di un processo automatico e su larga scala di conservazione attiva da parte di archivi discografici.

Il sistema HW/SW prototipato (*Photos of GHOSTS: Photos of Grooves and HOles, Supporting Tracks Separation*) possiede le seguenti caratteristiche:

- separa automaticamente le tracce presenti nel disco;
- riduce al minimo l'intervento dell'utente;
- utilizza HW *low-cost*;
- è robusto rispetto a polvere, graffi e ondulazioni anomale dei solchi;
- applica una curva di equalizzazione, opportunamente scelta (in funzione della data di incisione del disco, dell'etichetta discografica, ecc.) da una banca dati creata appositamente.

Il sistema impiega uno scanner modificato affinché mantenga, durante la ripresa fotografica, un allineamento ottimale della lampada rispetto al supporto (irradiamento della luce coassiale al disco) per tutti i solchi, senza utilizzare luce ultravioletta. Lo scanner usato (HP ScanJet 4890 Photo) è impostato con una profondità di 8 bit (a livelli di grigio) e con una risoluzione di 4800 dpi, senza correzione digitale.

## 2. Restauro

Sono stati progettati innovativi strumenti di restauro audio per le esigenze dell'Archivio Vicentini, apportando sensibili miglioramenti al filtro EMSR (*Ephraim-Malah Suppression Rule*), considerato uno dei riferimenti in questo campo [2], [8], [9]. Tale metodo impiega un'analisi del segnale mediante la trasformata di Fourier a breve termine e può essere pensato come un adattamento non stazionario del filtro di Wiener nel dominio della frequenza. Il guadagno dipende da due stime diverse del rapporto segnale/rumore (SNR):  $Y_{post}$ , che tiene conto dell'informazione del frame precedente, e  $Y_{prio}$ , corrispondente al tradizionale calcolo di SNR. Un parametro ( $\alpha$ ) controlla

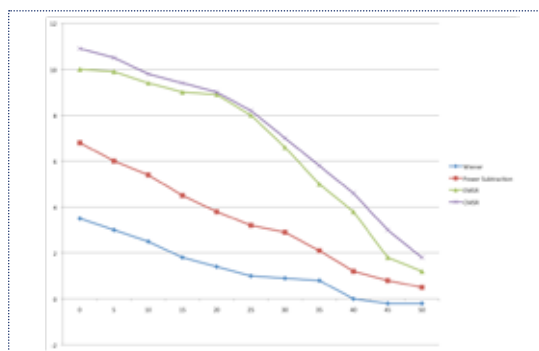
il bilanciamento tra l'informazione del frame corrente e quella del frame precedente. Tramite questo parametro si regola l'effetto di *smoothing* del filtro. Per evitare che, nel segnale restaurato, sia presente un disturbo costituito da numerose pseudo-sinusoidi di frequenza casuale, che nascono e muoiono in rapida successione (rumore musicale), si è progettato il filtro CMSR (*Canazza-Mian Suppression Rule*), basato su una stima di  $Y_{prio}$  su un numero variabile di frames. Analisi condotte su casi reali hanno evidenziato la capacità del filtro di ridurre i disturbi senza introdurre rumore musicale.

La figura 1 mostra il guadagno di CMSR in relazione al comportamento di alcuni tra i filtri più utilizzati nel campo del restauro audio: Filtro di Wiener, Power Subtraction, EMSR [2], [8].

## 3. Accesso

È stato definito un modello dei dati basato su XML e finalizzato alla catalogazione delle informazioni contenute nei documenti audio.

Per l'estrazione automatica dei metadati sono stati considerati strumenti sviluppati nel campo della *computer vision*. Per calcolare l'eccentricità del disco (che causa oscillazioni del *pitch* o distorsioni del suono) è stata sfruttata la letteratura prodotta nel campo dei sistemi per il riconoscimento dell'iride (*iris detection*) [10]. In particolare, il sistema progetta-



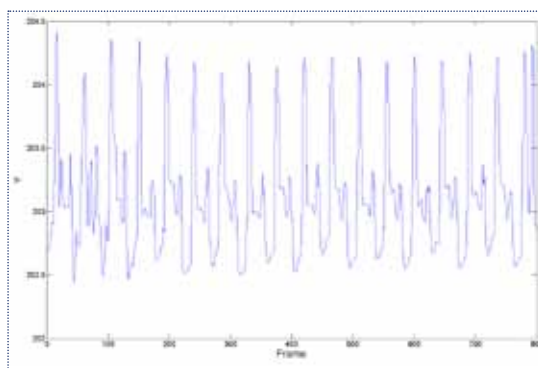
**Fig. 1** Guadagno introdotto dal filtro al variare del SNR d'ingresso (ascissa: SNR del segnale d'ingresso al filtro; ordinata: SNR d'uscita - SNR d'ingresso). Si può osservare il miglior comportamento di CMSR per tutti i rapporti SNR

to utilizza l'operatore integrodifferenziale sviluppato per il rilevamento dei confini della pupilla e dell'iride [11]. In questo contesto, viene utilizzato per estrarre i contorni del disco e quindi per rilevare la circonferenza del foro centrale. L'immagine statica viene acquisita garantendo il parallelismo tra il piano focale e il disco. Il sistema calcola quindi il raggio, il centro del disco e la deviazione tra questo e il centro del foro. Per separare automaticamente le tracce viene invece usato un rilevatore a soglia sulla funzione intensità luminosa, calcolata tra lo *specchio* e il margine esterno del disco fonografico.

Le deformazioni dei supporti fonografici (dischi *sbilanciati*) causano una sensibile variazione del *pitch* del segnale audio. Questo è dovuto al movimento ondulatorio a cui è soggetto il braccio del giradischi. Poiché le caratteristiche del moto oscillatorio possono essere messe in relazione con la variazione del *pitch*, esse costituiscono importanti metadati utili al restauro del segnale audio. In questo caso è stato utilizzato l'algoritmo di *features tracking* comunemente noto come Lucas-Kanade *tracker* [12]. L'algoritmo individua dei punti notevoli nell'immagine (*features*) che possono essere impiegati per confrontare due frame successivi e valutare gli spostamenti fra uno e l'altro. La tecnica, inizialmente concepita per l'allineamento di immagini, viene utilizzata in questo contesto nella sua implementazione come *features tracker*, che è in grado di tenere traccia degli spostamenti delle *features* da un frame al successivo (fig.2). Questi dati vengono memorizzati nella copia d'archivio per eventuali elaborazioni future (correzione del *pitch* sul segnale audio digitale).

#### 4. Allegati multimediali

File sonori e immagini generati con le metodologie descritte sopra sono reperibili in:  
[http://avires.dimi.uniud.it/tmp/DL/Experimental\\_Results.html](http://avires.dimi.uniud.it/tmp/DL/Experimental_Results.html)



**Fig.2** Evoluzione temporale della coordinata y di una *feature* di Lucas-Kanade posizionata sul braccio. Le evidenti oscillazioni indicano la presenza di un supporto imperfetto.

#### 5. Ringraziamenti

Questo lavoro è stato parzialmente supportato dal Joint Project “REstauro dell’archivio Vincentini di Verona e sua accessibilità come Audio e-Library” (2009-2010), finanziato dalla Fondazione Arena di Verona e dall’Università di Verona e svolto in collaborazione con Eye-Tech srl.

#### Riferimenti bibliografici

- [1] S. Canazza - Conservazione attiva e restauro audio dei 78 giri. Un caso di studio: Eternamente, in S. Canazza and M. Casadei Turroni Monti, M. (Eds.) - Ri-mediazione dei documenti sonori, Udine, Forum, 2006, p. 695-715
- [2] S. Canazza - Tecniche di filtraggio per il restauro audio: modelli a confronto, in S. Canazza and M. Casadei Turroni Monti, M. (Eds.) - Ri-mediazione dei documenti sonori, Udine, Forum, 2006, p. 259-338
- [3] S. Canazza - Note sulla conservazione attiva dei documenti sonori su disco, in Il suono riprodotto: storia, tecnica e cultura di una rivoluzione del Novecento, Torino, EDT, 2007, p. 87-107
- [4] S. Cavaglieri, O. Johnsen, F. Bapst - Optical Retrieval and Storage of Analog Sound Recordings, in A.E.S. 20th International Conference, Budapest, Hungary, 2001
- [5] V. Fadeyev, C. Haber - Reconstruction of

mechanically recorded sound by image processing, in *Journal of the Audio Engineering Society*, 12, 2003, p. 1172-1185

[6] S. Stotzer, O. Johnsen, F. Bapst, C. Sudan, R. Ingol - Phonographic Sound Extraction Using Image and Signal Processing, in *Proceedings of ICASSP*, 2004, p. 17-21

[7] S. Canazza, N. Orio - Digital Preservation and Access of Audio Heritage: A Case Study for Phonographic Discs, in *13th Conference on Digital Libraries*, 2009, Corfu, Greece, pp. 451-454

[8] S. Canazza - Noise and Representation Systems: A Comparison among Audio Restoration Algorithms. Lulu Enterprise, USA

[9] S. Canazza, A. Vidolin - Preserving Electroacoustic Music, In *Journal of New Music Research*. 30, 4, 2001, pp. 351-363

[10] J. Daugman - How iris recognition works, in *IEEE Transactions on circuits and systems for video technology*, 1, 2004, p. 21-30

[11] J. Daugman - New methods in iris recognition, in *IEEE Transactions on system, man, and cybernetics –part B: cybernetics*, 37, 5, 2007, p. 1167-1175

[12] J. Shi, C. Tomasi - Good Features to Track, in *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WS, 1994, p. 593-600

# Progetto di un sistema per la conservazione a lungo termine e a norma di legge di documenti elettronici

Raffaele Conte<sup>1,2</sup>, Andrea Renieri<sup>2</sup>

<sup>1</sup>CNR, Istituto di Fisiologia Clinica, <sup>2</sup>QualiMedLab s.r.l.



**Abstract.** Nell'articolo viene presentato il lavoro svolto dall'Istituto di Fisiologia Clinica del CNR per la progettazione di un sistema di gestione elettronica dei dati prodotti durante l'attività clinica.

Per garantire una conservazione di lungo periodo e il rispetto dei termini di validità legale, è stato necessario utilizzare strumenti e metodi di prevenzione nei confronti dell'invecchiamento dei supporti, dell'obsolescenza dei formati software e della scadenza delle firme digitali. Il sistema è stato progettato rispettando la normativa CNIPA e in conformità con lo standard ISO OAIS. Il risultato finale è stato la realizzazione di un sistema molto semplice dal punto di vista architetturale, modulare e flessibile in vista di un'esportazione verso altri enti. Particolare attenzione è stata posta agli strumenti di indicizzazione e all'utilizzo di software open source.

## 1. Contesto

L'Istituto di Fisiologia Clinica del CNR ha sviluppato negli anni un sistema per la gestione elettronica dei dati prodotti durante l'attività clinica. In accordo con il processo di dematerializzazione, avviato in Italia già da qualche anno, è possibile velocizzare i processi clinico-amministrativi minimizzando la produzione di documentazione cartacea (o di altri supporti analogici, ad esempio lastre) con conseguenti benefici non solo economici. Tali vantaggi vengono però vanificati nei casi in cui i documenti prodotti richiedano la conservazione, per un certo periodo di tempo, per fini fiscali e/o amministrativi. Ad oggi, infatti, anche i documenti elettronici corredati di firma digitale possono perdere la loro validità legale nel tempo se non conservati secondo le regole dettate dal CNIPA e dai Ministeri competenti in materia. Le procedure per la conservazione di documenti elettronici, se implementate correttamente, consentono altresì la conservazione di documenti analogici convertiti in formato elettronico.

## 2. Obiettivi

La progettazione del sistema ha tenuto in con-

siderazione la necessità di predisporre metodiche utili al mantenimento, inalterato nel tempo, dei documenti conservati, al fine di prevenire l'obsolescenza dei supporti di memorizzazione, il superamento dei formati software dei documenti e la scadenza delle firme digitali. Inoltre, era necessario implementare strumenti di indicizzazione dei documenti e funzioni per il reperimento e l'esibizione degli stessi nel tempo.

Un requisito fondamentale riguardava inoltre la possibilità di erogare il servizio di conservazione non solo verso l'interno, ma anche verso soggetti esterni (ad esempio, le Aziende Sanitarie).

## 3. Strumenti e Metodi

Il sistema per l'archiviazione e la conservazione dei documenti informatici è stato progettato tenendo conto della normativa CNIPA [2], [3] ed in conformità con lo standard ISO OAIS (*Open Archive Information System*) [4]. Tale standard è frutto di uno studio promosso dalla Consultative Committee for Space Data System che ha elaborato, nel 2002, un modello di sistema per la conservazione a lungo termine dei documenti. Tale modello, applicabile ad ogni tipologia di archivio, sia car-

taceo sia elettronico, propone degli schemi organizzativi che facilitano la strutturazione del sistema e la gestione delle problematiche derivanti dall'attività di archiviazione e conservazione a lungo termine. Lo standard OAIS è stato preso in considerazione anche nei tavoli tecnici interministeriali [5] promossi dal CNIPA nel 2007 al fine di chiarire la normativa riguardante la conservazione digitale. L'OAIS rappresenta sicuramente un trattato fondamentale per quanto riguarda la conservazione di documenti in un sistema di archiviazione digitale, tanto da essere diventato standard (ISO 14721).

Precedentemente alla fase di progettazione, è stato necessario sia uno studio dettagliato dei possibili supporti di memorizzazione adatti allo scopo, sia una rendicontazione delle possibili tipologie di documento che potevano afferire al sistema. In prima istanza, si è fatto riferimento a tipologie di documento di carattere principalmente clinico-sanitario, per cui si è deciso di far riferimento allo standard CDA (*Clinical Document Architecture*) definito dall'organizzazione HL7 [6]. In questo modo è stato possibile definire metadati ed attributi, con denominazione, sintassi e semantica già previste da tale standard.

L'intera procedura di conservazione di un documento può essere scomposta nelle seguenti azioni:

1. ricezione del documento;
  2. verifica dell'integrità del documento e della validità della firma associata;
  3. indicizzazione del documento tramite i metadati nel database degli indici;
  4. inserimento del documento in un volume di conservazione ed inserimento della sua impronta digitale (*hash*) in un file indice associato al volume;
  5. chiusura del volume, al suo completamento, mediante apposizione della firma digitale;
  6. archiviazione del volume e dell'indice associato su un supporto di memorizzazione.
- Un volume di conservazione è costituito da

un insieme omogeneo “per qualche aspetto” di documenti acquisiti dal sistema. Il volume può trovarsi in diversi stati:

- *in attesa di conservazione*, ed in questo caso è idoneo ad accogliere ulteriori documenti;
  - *chiuso*, a seguito dell'operazione di estrazione ed inserimento degli *hash* nell'indice, firmato e posto all'interno del volume stesso;
  - *conservato*, quando un volume chiuso viene trasferito sui supporti di memorizzazione.
- Si è deciso di compilare l'indice utilizzando il formato XML e di inserire, per ogni documento, oltre all'*hash*, anche i metadati e gli attributi descrittivi dei documenti. In questo modo il volume diventa completamente autoconsistente ed indipendente dal sistema di conservazione che l'ha creato.

Il sistema è stato progettato con l'obiettivo di essere semplice e modulare. Tali caratteristiche sono da considerarsi essenziali per un sistema che deve garantire la conservazione dei documenti a lungo termine. Esso non deve essere statico ed immutabile, bensì deve potersi evolvere e riconfigurare nel tempo, senza essere mai stravolto nella sua struttura basilare. Al contrario, le architetture proprietarie, oltre che costose, sono spesso “monolitiche” ovvero difficilmente in grado di evolvere o modificarsi, se non con aggiornamenti forniti dagli stessi produttori. L'architettura, basata sulla tecnologia dei *web service*, risulta flessibile potendo accogliere e definire diversi blocchi architetturali integrando, quando opportuno, software open source esistenti.

Il sistema di conservazione si compone quindi di tre moduli che interagiscono fra loro (*fig.1*):

- Mediatore;
- Archivio;
- Interfaccia.

Il Modulo Mediatore ed il Modulo Archivio sono i gestori dell'archiviazione e della conservazione dei documenti. Il primo deve poter mettere a disposizione funzioni e metodi uti-

li per l'interazione dei client con il sistema di conservazione, in particolare per il passaggio dei documenti. Il sistema prevede che i client possano dialogare con il Modulo Mediatore per mezzo dei Moduli Interfaccia. Esistono più Moduli Interfaccia (uno per ogni client) che dialogano da un lato con il sistema di gestione documentale dell'organizzazione che invia i documenti, dall'altro con il Modulo Mediatore del sistema di conservazione tramite protocollo SOAP (fig.2). Il Modulo Archivio è composto dagli strumenti di *storage* e dalla base di dati utile per l'indicizzazione dei documenti.

Il sistema riprende alcuni concetti proposti dal CNIPA nell'architettura SPCoop, per la cooperazione tra pubbliche amministrazioni [7]. È stato realizzato un prototipo del sistema descritto, dove il Modulo Mediatore è stato implementato utilizzando il framework Axis2 dell'Apache Software Foundation. Il Modulo Archivio fa uso di un *disk array* e nastri LTO (tecnologia valutata come la più adatta per la conservazione a lungo termine di documenti).

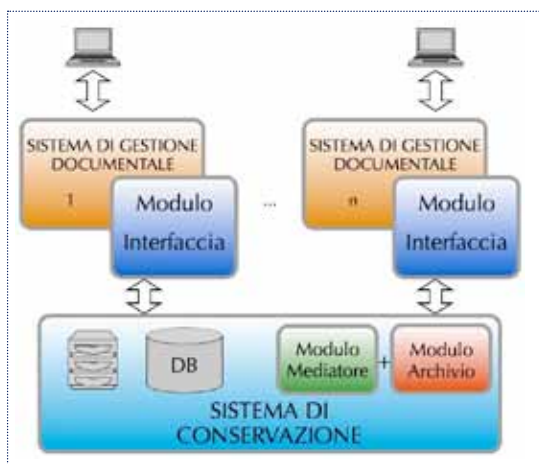


Fig. 2 Architettura del sistema

Un Modulo Interfaccia di test è stato realizzato in PHP e implementa un caso particolare

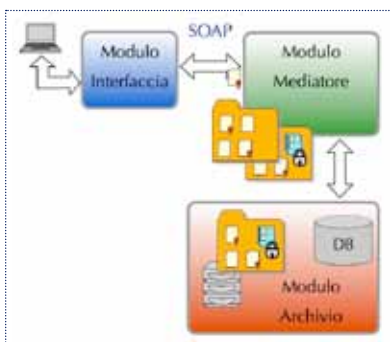


Fig. 1 Relazioni tra i moduli funzionali

per l'inserimento dei documenti via web.

#### 4. Risultati

Scopo del progetto è stato la realizzazione di un sistema, conforme alla legge italiana, per l'archiviazione e la conservazione a lungo termine di documenti elettronici. In fase di implementazione si è cercato di integrare nel sistema, dove possibile,

progetti open source già esistenti, dimostrando che un'architettura complessa può basarsi anche su software aperti. In particolare, alcune funzioni di verifica e controllo (raggiungimento della soglia di chiusura dei volumi, verifica periodica dei supporti di memorizzazione, ecc.) sono state affidate al software Nagios. L'architettura è stata appositamente progettata per offrire i servizi a sistemi documentali di altre organizzazioni, con moduli esterni al sistema, implementati da terzi, ma capaci di interfacciarsi ed utilizzare le funzionalità messe a disposizione dal servizio di conservazione.

Fra i risultati ottenuti vi è la dimostrazione che il problema della conservazione a lungo termine di documenti elettronici, apparentemente complesso, può essere risolto tramite la realizzazione di un sistema che ha come sua caratteristica essenziale proprio la semplicità architetturale, la quale ne diventa, al contrario, un requisito essenziale.

#### Riferimenti bibliografici

[1] Renieri A, Progetto di un sistema a norma di legge per la conservazione a lungo termine di documenti informatici, Tesi di Laurea Specialistica in Informatica, Facoltà di Scienze Matematiche Fisiche e Naturali, Università di Pisa, Anno Accademico 2007-2008

[2] Deliberazione CNIPA n. 11/2004, Regole tecniche per la riproduzione e conservazione di documenti su supporto ottico idoneo a ga-

rantire la conformità dei documenti agli originali - articolo 6, commi 1 e 2, del Testo unico delle disposizioni legislative e regolamentari in materia di documentazione amministrativa, di cui al decreto del Presidente della Repubblica 28 dicembre 2000, n. 445

[3] Proposta Deliberazione CNIPA, Regole Tecniche in materia di formazione e conservazione di documenti informatici, gennaio 2008

[4] Consultative Committee on Space Data System, Reference Model for an Open Archival Information System (OAIS), CCSDS Secretariat, National Aeronautics and Space Administration, gennaio 2002

[5] CNIPA, La dematerializzazione della documentazione amministrativa, Libro Bianco del Gruppo di Lavoro interministeriale per la dematerializzazione della documentazione tramite supporto digitale, CNIPA, marzo 2006

[6] Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, Shabo A, HL7 Clinical Document Architecture, Release 2, American Medical Informatics Association, ottobre 2005

[7] CNIPA, Sistema Pubblico di Connettività (SPC), <http://www.cnipa.gov.it/>, CNIPA, 2005



# Digital libraries su Grid per preservare e diffondere il patrimonio umanistico

Antonio Calanducci<sup>1</sup>, Jorge Sevilla<sup>1</sup>, Monica Saso<sup>2</sup>, Alessandro De Filippo<sup>2</sup>, Stefania Iannizzotto<sup>2</sup>, Roberto Barbera<sup>2,1,3</sup>, Giuseppe Andronico<sup>1,3</sup>, Francesco De Mattia<sup>4,5</sup>, Domenico Vicinanza<sup>6</sup>

<sup>1</sup>INFN Sezione di Catania, <sup>2</sup>Università degli Studi di Catania, <sup>3</sup>Consorzio COMETA, <sup>4</sup>Conservatorio di Musica "A. Boito" di Parma, <sup>5</sup>ASTRA Project, <sup>6</sup>Delivery Advanced Network Technology to Europe, UK



**Abstract.** L'articolo descrive l'utilizzo delle Grid nella digitalizzazione del patrimonio letterario cartaceo. L'elevata qualità delle scansioni comporta, infatti, una dimensione considerevole degli archivi digitali. Le *Data Grid* offrono enormi capacità di storage, garantendo un luogo sicuro e ideale per la conservazione a lungo termine. I meccanismi di autenticazione e di autorizzazione di una Grid permettono di fornire un accesso controllato agli archivi per singoli utenti, gruppi o intere comunità. I servizi per i metadati, inoltre, consentono di organizzare i contenuti per effettuare ricerche efficienti. In questo articolo vengono presentati due esempi di digital library: gli archivi dei manoscritti dello scrittore Federico De Roberto e gli archivi del progetto "Civiltà Musicale Napoletana". Un prototipo funzionante del repository di De Roberto è stato implementato su Grid usando gLibrary, un sistema sviluppato dalla Sezione di Catania dell'INFN sull'infrastruttura Grid siciliana del Consorzio COMETA.

## 1. Introduzione

Negli ultimi decenni, la rivoluzione digitale ha profondamente modificato le tradizionali metodologie di conservazione e diffusione del patrimonio culturale. Lo sviluppo tecnologico digitale ha migliorato notevolmente la qualità dei servizi di tutela e di condivisione offerti dalle biblioteche, dagli archivi e dai musei. In particolare, la pratica di digitalizzazione permette di sottrarre il materiale librario ai fenomeni di deterioramento dovuti alla manipolazione, fornisce una riproduzione soddisfacentemente fedele dell'originale e garantisce la fruibilità dei contenuti. I problemi relativi alla conservazione riguardano, una volta digitalizzato, anche l'oggetto digitale di cui deve essere garantita la durata nel tempo. Inoltre, la digitalizzazione del patrimonio culturale attraverso la riproduzione dei dati ne consente una diffusione ed una divulgazione su ampia scala.

## 2. Digitalizzazione e conservazione dei dati

La quantità di dati digitali prodotti dipende

dalla tipologia e dalla mole dei documenti e dal numero di postazioni di lavoro operative: un sistema di digitalizzazione che lavori in parallelo su più supporti può generare quotidianamente centinaia di Gigabyte. Queste notevoli quantità di dati pongono il problema di una loro archiviazione e conservazione sicura. La conservazione e l'accesso alle informazioni digitali sono garantiti da un sistema di *storage* affidabile e capace di prevenire qualunque tipo di perdita dei dati acquisiti. In particolare, è necessario garantire la creazione di più copie dei dati digitali da archiviare preferibilmente su supporti conservati in luoghi diversi (*immagazzinamento ridondante*).

La rapida evoluzione delle tecnologie informatiche comporta un elevato grado di obsolescenza dei prodotti elettronici e pertanto bisogna trovare soluzioni capaci di garantire la conservazione del patrimonio culturale digitalizzato e di permetterne l'accesso nel tempo. In particolare, i rischi riguardano l'"invecchiamento precoce" dei formati elettronici, la



transitorietà dei supporti fisici di archiviazione e la caducità dei dispositivi hardware. La finalità di diffusione e divulgazione è favorita da una fruizione immediata dei materiali culturali digitali.

L'elemento cardine di questo processo è un'interfaccia utente intuitiva ed *"easy to use"*, che garantisca un accesso semplice e rapido ai dati e consenta una ricerca efficace grazie a una serie di parole chiave ed un set di metadati standard, composti da informazioni sui documenti, integrate con i file.

### 3. Casi d'uso

Possono essere considerati due casi d'uso per dimostrare come le *digital libraries* su Grid possano garantire una duratura conservazione del patrimonio culturale umanistico: l'archivio dei lavori dello scrittore siciliano Federico De Roberto (1861-1927), che raccoglie quasi 8000 scansioni, e l'archivio musicale del progetto Civiltà Musicale Napoletana, composto da più di 250.000 documenti digitali.

#### 3.1 Il fondo Federico De Roberto

All'interno del fondo manoscritto di De Roberto, custodito presso la Biblioteca della Società di Storia Patria per la Sicilia Orientale di Catania, sono conservati manoscritti, dattiloscritti, bozze di stampa, articoli di giornale, carteggi, carte geografiche, ritagli, appunti e fotografie, per un totale di circa 10.000 documenti.

La maggior parte dei documenti sono di dimensioni variabili dal formato A5 a quello A3, insieme a documenti di formati particolari (7x7 cm, 3x15cm ecc). Molti fogli hanno aggiunte e note laterali incollate a margine e ripiegate, che determinano una dimensione totale di superficie superiore

al formato A3. I documenti riportano numerosi interventi di correzione, tagli al testo delle opere e segni di rimando.

Oltre al set di metadati acquisito in forma automatica dal dispositivo di scansione (*ImageWidth, ImageHeight, XResolution, FileSize, CreationDate* e *ModifyDate*), tutti i metadati sono stati inseriti attraverso la specifica XMP (*eXtensible Metadata Platform*), un formato di gestione di tipo XML nei software Adobe che rispetta lo standard *Dublin Core*. L'inserimento è avvenuto in batch con l'ausilio del software Adobe Photoshop CS.

Il set di metadati comprende *Document, Title, Author, Description, Description Writer, Keywords, Copyright Status* e *Copyright Notice*. Le immagini esportate sono state lavorate fino alla creazione di file PDF in grado di ordinarle e organizzarle in pagine, comprimendole e ottimizzandole per la fruizione online.

#### 3.2 Gli spartiti della Civiltà Musicale Napoletana

Gli spartiti della Civiltà Musicale Napoletana sono stati progressivamente digitalizzati all'interno di un *repository* contenente adesso più di 250.000 documenti. Le partiture sono copie di esecuzione, attualmente archiviate presso le librerie musicali di tutto il mondo. La maggioranza di esse, oltre che in Italia, è conservata a Londra, Parigi, Dresda e San Pietroburgo.

La digitalizzazione ha consentito di raccogliere in un unico archivio lavori, documen-

ti e manoscritti appartenenti a diverse collezioni: la biblioteca del Tesoro di San Gennaro, del Conservatorio di Napoli San Pietro a Maiella, la libreria Gerolomini ed il Fondo Gustavo Alfonso Nosedà



Fig. 1 Scansione di una *Serenata* inedita di Leonardo Leo per "violini, viola, canto e b.c."

del Conservatorio di Milano.

#### 4. La Piattaforma gLibrary

gLibrary rappresenta una soluzione sicura ed efficace per conservare e condividere file attraverso un'infrastruttura Grid. È uno strumento flessibile che consente di gestire qualsiasi tipo di documento digitale, anche di grandi dimensioni. L'interfaccia di *browsing*, che trae ispirazione da iTunes di Apple, il software che consente di organizzare le librerie multimediali di iPod/iPhone, è molto semplice da usare e consente un'agevole ricerca e download dei file desiderati, attraverso un front-end web intuitivo ed interattivo, sempre accessibile. Gli utenti non devono preoccuparsi della complessità del sistema Grid su cui si basa l'intera infrastruttura, fornendo un accesso trasparente agli *storage* Grid come se fosse un grande disco virtuale.

gLibrary consente di memorizzare, organizzare, cercare e recuperare *digital asset* in una piattaforma Grid: di conseguenza può essere utile per diversi tipi di utenti che necessitano di un metodo sicuro per conservare e condividere dati di qualsiasi tipo. Possono inoltre essere definiti permessi e autorizzazioni per accedere alle librerie ospitate sul server, e quindi sugli *asset* stessi, al fine di concedere o restringere l'accesso a singoli o gruppi di utenti, sfruttando il sistema di autenticazione e autorizzazione offerto dal middleware Grid.

I dati sono salvati su Grid e possono essere criptati e replicati su diversi *storage element*, garantendo al contempo massima sicurezza ed alta disponibilità.

Tutti gli *asset* organizzati in una libreria di gLibrary sono organizzati per *tipo*: un elenco di attributi specifici per descrivere ogni *tipo* di risorsa che viene definita dal fornitore del *repository*. Questi attributi possono essere utilizzati da parte degli utenti per le opera-

zioni di ricerca. Inoltre possono essere definiti dei sotto-tipi che specializzano uno esistente. I file registrati in una libreria sono associati alla corretta tipologia durante il processo di registrazione/upload.

I server di gLibrary possono ospitare più librerie, con diversi *tipi* di struttura gerarchica e accessibili da utenti diversi. Inoltre, i *tipi* e le relative voci sono visibili solo da utenti con privilegi appropriati. Gli *asset* possono anche essere organizzati in *collection*, che consentono di raggruppare elementi appartenenti a *tipi* diversi ma che hanno una qualche relazione.

Ogni *asset*, *tipo* e *collection* possiede delle ACL (*Access Control List*) che ne restringono l'utilizzo, permettendo ai proprietari degli oggetti digitali di concedere l'accesso a un'intera organizzazione virtuale, a gruppi di utenti o a un singolo utente. Gli utenti che non dispongono delle corrette autorizzazioni, non sono in grado di visualizzare attraverso l'interfaccia di navigazione gli elementi ai quali è stata applicata la restrizione.

#### 5. Il middleware Grid e gLibrary

gLibrary è costruito sui servizi del middleware gLite. Nato dallo sforzo di più di ottanta persone in dodici diversi centri di ricerca accademici e industriali, come parte del progetto EGEE, gLite fornisce un framework per il *porting* di applicazioni su Grid, che permette di sfruttare in modo ottimale risorse etero-

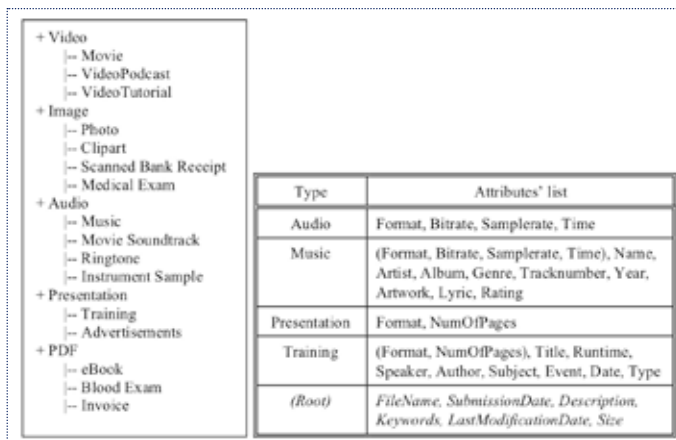


Fig. 2 Esempio di un albero dei *tipi* e della relativa lista di attributi

genee di calcolo e di *storage* geograficamente distribuite.

gLibrary, essendo un'applicazione *data-oriented*, fa uso dei seguenti servizi di gLite:

- *Storage Element (SE)*: forniscono un accesso uniforme alle risorse di archiviazione dati. Possono dare accesso a dischi singoli, *array* di dischi o sistemi di archiviazione su nastro;
- *AMGA Metadata Catalog*: archivia i metadati descrivendo i contenuti dei file su Grid, permettendo agli utenti una ricerca basata sulle loro descrizioni;
- *LCG File Catalog (LFC)*: tiene traccia dei filename logici associandoli alle locazioni fisiche delle repliche di un file archiviato su uno o più *Storage Element*;
- *Virtual Organization Membership Service (VOMS)*: un servizio che permette la definizione dettagliata dei privilegi e dei ruoli degli utenti, secondo entità astratte chiamate *Virtual Organization*, che raggruppano utenti, istituzioni e risorse all'interno dei loro rispettivi domini amministrativi;
- *Information service*: fornisce informazioni riguardo le risorse Grid e il loro status; in particolare è utilizzato per scoprire gli *Storage Element* disponibili per una data *Virtual Organization*.

## 6. L'implementazione del *De Roberto digital repository* con gLibrary

Una volta portate a compimento le fasi vere e proprie di acquisizione, si è posto il problema di trovare un metodo per rendere disponibili online le immagini digitalizzate del fondo De Roberto, in modo da garantirne l'accesso secondo una predefinita strategia di fruizione. Si è ritenuto opportuno trovare un sistema in gra-

do di garantire la disponibilità continua in rete dei documenti e la consultazione da ogni parte del mondo attraverso un'interfaccia semplice e intuitiva.

La flessibilità e l'estensibilità offerti dall'infrastruttura Grid e da gLibrary hanno consentito una perfetta integrazione nell'archiviazio-

ne e nella catalogazione in rete dei manoscritti digitalizzati del fondo.

Adottando le risorse Grid offerte dal Consorzio COMETA, è stato così realizzato il *De Roberto digital repository* (DRdr), di cui è disponibile una versione beta all'indi-

rizzo: <https://glibrary.ct.infn.it>.

Attraverso l'interfaccia per l'upload di gLibrary, che consente di selezionare file multipli, le immagini digitali ad altissima risoluzione dei manoscritti ed i file PDF a bassa risoluzione, sono stati caricati in automatico sugli *Storage Element* dell'infrastruttura Grid di COMETA.

Durante il processo di upload sono stati estratti automaticamente i metadati già inseriti nell'header delle scansioni, salvandoli sul server per i metadati AMGA.

Non appena l'utente effettua il login dalla homepage di gLibrary, utilizzando il suo certificato personale di tipo X.509, gli è immediatamente possibile interagire con l'interfaccia di navigazione. Il *De Roberto digital repository* consente una ricerca agevole tra gli *asset* rappresentati dai manoscritti digitalizzati: i contenuti sono organizzati per caratteristiche fisiche e semantiche attraverso la definizione di tipi e categorie e i risultati sono filtrati attraverso la selezione di uno o più attributi a scelta da liste predefinite.

I filtri, definiti per ogni tipo, si dividono in: *DocumentGenre*, il genere a cui appar-

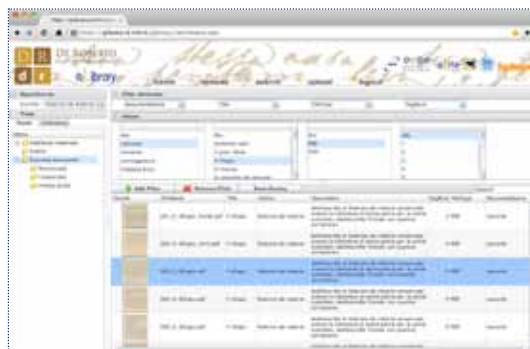


Fig. 3 Esempio di navigazione attraverso l'utilizzo dei filtri per la ricerca del dattiloscritto *Il Rifugio* di Federico De Roberto

tiene l'opera (novella, saggio, tragedia lirica ecc.); *Title*, il titolo dell'opera; *FileType*, il formato del file (TIFF, PDF, ecc); *ScanQuality*, la qualità della scansione; *DocumentType*, il tipo di documento (manoscritto, dattiloscritto, bozza di stampa ecc.); *PublicationStatus*, lo status di pubblicazione (pubblicato, inedito ecc.); *PublicationYear*, l'anno di pubblicazione; *Publisher*, il nome dell'editore; *Location*, l'ubicazione fisica del documento all'interno del fondo, rappresentato dal numero della busta entro il quale è contenuto.

La selezione dei filtri, realizzata a cascata, permette all'utente di muoversi agevolmente attraverso il processo di ricerca. La selezione del valore di un filtro influenza dinamicamente la lista dei valori dei filtri successivi, consentendo di raggiungere rapidamente e in maniera razionale il risultato della ricerca.

Consideriamo qualche esempio nell'utilizzo dei filtri per raggiungere un documento specifico o una serie di documenti.

Primo esempio. Uno studioso può avere bisogno di cercare, tra i documenti digitalizzati del fondo De Roberto, tutte le bozze di stampa dell'anno 1919. In questo caso, selezionerà dall'albero dei *tipi* disponibili "*Scanned Documents*", poi "*testi a stampa*". La selezione del filtro "*PublicationYear*", con il relativo valore "*1919*" tra gli anni disponibili, produrrà come risultato l'elenco di tutti documenti che soddisfano queste caratteristiche. La ricerca può essere ulteriormente affinata selezionando l'ulteriore filtro "*Publisher*", che consentirà di visualizzare le bozze di stampa pubblicate per editore.

Secondo esempio. Un linguista può aver

bisogno di intraprendere lo studio e l'analisi delle varianti d'autore dell'"Ermanno Raeli" di Federico De Roberto. Dovrà cercare in questo caso tutte le versioni disponibili, dal manoscritto all'opera edita. Attraverso la selezione della voce "*Scanned Documents*", tra i *tipi* e l'impostazione del



Fig. 4 Prima pagina del manoscritto di appunti per la stesura del romanzo *I Viceré* di Federico De Roberto

filtro "*Title*" con il relativo valore "*Ermanno Raeli*" tra i titoli delle opere disponibili, sarà in grado di visualizzare l'elenco dei manoscritti e di tutte le successive versioni delle bozze di stampa, con le relative modifiche e correzioni, dell'opera presa in considerazione, ordinabili per anno.

Il sistema, così organizzato, consente di effettuare ricerche su tutte le informazioni a livello contenutistico, ma anche di cercare dettagli sulla conservazione fisica dei documenti.

È possibile ottenere risultati anche sulla qualità delle scansioni e sui formati fisici dei documenti e a una serie dettagliata di informazioni aggiuntive attraverso un semplice click sul singolo risultato, che consente di visualizzare, attraverso una finestra pop-up, i metadati associati ai singoli file, inseriti in fase di digitalizzazione.

Il download del documento desiderato può essere avviato cliccando su uno dei link disponibili rappresentati all'interno della lista delle repliche. L'archiviazione attraverso gli *Storage Element* della piattaforma Grid garantisce la custodia dei dati, attraverso la conservazione di copie molteplici dislocate su server in località geografiche differenti, e la sicurezza nella conservazione, grazie all'affidabilità dei sistemi di *storage* e l'archiviazione ridondante.

## 7. Conclusioni

Con l'implementazione e lo sviluppo del *De Roberto* digital repository su una infrastruttura Grid attraverso la piattaforma gLibrary è stato possibile dimostrare come le scienze umanistiche possano sfruttare le nuove tecnologie per la conservazione e la disponibilità immediata dei contenuti, che non necessita di alcuna conoscenza tecnica specifica.

La salvaguardia del patrimonio culturale può essere garantita attraverso l'attività di digitalizzazione degli oggetti culturali, mentre l'accesso ai documenti da parte della comunità umanistica è reso possibile grazie a gLibrary e all'infrastruttura Grid. È un esempio di come le Scienze Umanistiche, dotandosi di strumenti estremamente avanzati in grado di potenziare i propri ambiti di ricerca, possano aprirsi a collaborazioni interdisciplinari che, come questa esperienza dimostra, possono produrre risultati sorprendenti.

## Riferimenti bibliografici

[1] A. Calanducci, R. Barbera, A. De Filippo, F. De Mattia, S. Iannizzotto, M. Saso, J. Sevil-la, D. Vicinanza, Digital Libraries on the Grid

to preserve cultural heritage in Proceedings of the First EELA-2 Conference, B. Marechal et al. (Eds.), CIEMAT 2008.

<https://glibrary.ct.infn.it/t/EELA-3-paper-final.doc>

[2] A. Calanducci, C. Cherubino, L. N. Ciuffo, D. Scardaci, A Digital Library Management System for the Grid in Fourth International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2007) at 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE-2007), GET/INT Paris, France, June 18-20, 2007

[https://glibrary.ct.infn.it/glibrary/downloads/gLibrary\\_paper\\_v2.pdf](https://glibrary.ct.infn.it/glibrary/downloads/gLibrary_paper_v2.pdf)

## Media

Versione beta del *De Roberto* Digital Repository: <https://glibrary.ct.infn.it>

## Immagini:

[https://glibrary.ct.infn.it/t/garr\\_immagini.pdf](https://glibrary.ct.infn.it/t/garr_immagini.pdf)

[http://glibrary.ct.infn.it/t/deroberto/garr\\_immagini2.pdf](http://glibrary.ct.infn.it/t/deroberto/garr_immagini2.pdf)

## Video:

[www.youtube.com/watch?v=r0zX0sQS5ag](http://www.youtube.com/watch?v=r0zX0sQS5ag)



# Identificativi Persistenti: jNBN, un'applicazione JEE per la gestione di un'infrastruttura nazionale NBN

Emanuele Bellini<sup>1</sup>, Mario Incarnato<sup>2</sup>, Roberto Puccinelli<sup>2</sup>, Marco Spasiano<sup>2</sup>

<sup>1</sup>Fondazione Rinascimento Digitale, <sup>2</sup>Consiglio Nazionale delle Ricerche



**Abstract.** L'utilizzo dei cosiddetti Identificativi Persistenti permette la creazione di una infrastruttura modulare, aperta e scalabile per l'identificazione e l'accesso di risorse a livello nazionale e internazionale e rappresenta quindi una caratteristica fondamentale dei sistemi di digital library. Nell'ambito delle biblioteche è stato definito il National Bibliography, basato sullo standard URN e attualmente promosso dalla Conference of European National Librarians (CENL). L'articolo descrive la genesi e le funzionalità del software jNBN, il cui scopo è quello di gestire lo sviluppo di un registro/resolver NBN (National Bibliography Number) italiano.

jNBN è stato testato con successo su una infrastruttura pilota nazionale, che coinvolge la Biblioteca Nazionale Centrale di Firenze, il CNR, la Fondazione Rinascimento Digitale e l'Università di Milano, tutti partner dell'iniziativa NBN italiana. Il prossimo passo sarà quello di partecipare alla realizzazione di un *global resolver* NBN di livello europeo.

## 1. Introduzione

La disponibilità di riferimenti stabili e certificati per le risorse su internet è fondamentale per le applicazioni di digital library, non solo per identificare dette risorse in modo affidabile, ma anche per garantirne l'accessibilità nel tempo. Iniziative di livello internazionale, quali l'European Digital Library [1] ed Europea [2], dimostrano l'importanza di questi temi.

Una possibile soluzione è quella di associare ad ogni risorsa digitale di interesse un Identificatore Persistente (*Persistent Identifier*, PI) in grado di certificare la sua autenticità e assicurarne l'accessibilità nel lungo periodo.

Attualmente sono disponibili alcune soluzioni sia a livello di standard per la definizione degli identificatori che di tecnologie per l'implementazione della infrastruttura di gestione [24], ma lo scenario presente mostra che non è verosimile imporre una sola tecnologia o un solo registro per tutte le risorse informative e documentarie diffuse via web. Le modalità identificative e la granularità di dette risorse cambiano a seconda delle finalità del-

le diverse comunità di produttori, gestori ed utilizzatori.<sup>1</sup>

Esempi di tali soluzioni sono l'Handle System [12], [26], [27], Il Document Object Identifier (DOI) [11], l'Archival Resource Key (ARK) [10], il Persistent URL (PURL) [13] ed il Library of Congress Control Number (LCCN) [14]. Per maggiori dettagli si veda [29].

Nell'ambito delle biblioteche è stato definito il National Bibliography Number (NBN-RFC 3188), basato sullo standard URN (RFC 2141) e correntemente promosso dalla Conference of European National Librarians (CENL). Il formato di questi identificativi prevede che l'insieme delle risorse digitali sia suddiviso in domini gerarchicamente ordinati, facenti riferimento, a livello internazionale, ad unico registro mantenuto dalla Library of Congress. Ogni Stato è responsabile dell'im-

<sup>1</sup> Da diversi anni, queste comunità hanno avvertito l'esigenza di realizzare un internet governance ed un'infrastruttura tecnologica in grado di assicurare una reale ed effettiva interoperabilità tra i diversi sistemi di produzione, gestione, identificazione e certificazione delle risorse informative, che consenta una condivisione consapevole delle stesse risorse ed un'integrazione razionale dei relativi servizi.

plementazione del proprio registro e del proprio servizio di risoluzione nazionale. Al vertice di ogni sistema nazionale c'è il rispettivo *top level domain* (NBN:IT per l'Italia), gestito dalla relativa biblioteca nazionale. Ciascun dominio nazionale è a sua volta suddiviso in sotto-domini, sull'esempio del sistema DNS. Ciascun identificativo NBN relativo ad una particolare risorsa è composto dal nome del dominio di appartenenza seguito da un codice alfanumerico, il cui formato è stabilito dalle policy dei singoli sistemi nazionali.

Esula dagli scopi del presente articolo una discussione dettagliata sulle motivazioni per le quali chi scrive ritiene che NBN sia una valida soluzione per la realizzazione di un'infrastruttura internazionale di gestione di identificativi persistenti. È sufficiente dire in questa sede che esso costituisce uno standard aperto, utilizzabile gratuitamente da qualunque organizzazione, e supporta la distribuzione delle responsabilità, consentendo a ciascun soggetto partecipante di mantenere il controllo completo sull'assegnazione degli identificativi alle proprie risorse.

## 2. L'iniziativa NBN in Italia

Il progetto per lo sviluppo di un registro/*resolver* NBN italiano è iniziato nel 2007 come collaborazione tra Fondazione Rinascimento Digitale, la Biblioteca Nazionale Centrale di Firenze, l'Università degli Studi di Milano ed il Consorzio Interuniversitario Lombardo per l'Elaborazione Automatica (CILEA). Dopo un anno è stato rilasciato un primo prototipo con lo scopo di dimostrare la validità dell'approccio gerarchico. Il prototipo utilizzava alcune funzionalità di DSpace ed Arc ed esponeva all'utente un'interfaccia web molto essenziale, basata su PHP. La gerarchia dei domini era limitata ad un massimo di due livelli.

La seconda ed attuale fase dell'iniziativa NBN italiana è basata su una partnership diversa, che coinvolge il Consiglio Nazionale

delle Ricerche (CNR), la Biblioteca Nazionale Centrale di Firenze, la Biblioteca Nazionale Centrale di Roma, l'Istituto Centrale per il Catalogo Unico (ICCU), la Fondazione Rinascimento Digitale, l'Università degli Studi di Milano e l'Agenzia Spaziale Italiana (ASI). Tali soggetti hanno costituito un organo di governo per il coordinamento delle attività di sperimentazione e la definizione delle policy.

All'inizio del 2009 è stato rilasciato il software jNBN, sviluppato dal Consiglio Nazionale delle Ricerche con il contributo di Fondazione Rinascimento Digitale e basato sul Framework JADA di proprietà del CNR. Tale software, realizzato su piattaforma Java Enterprise, ha eliminato la necessità di utilizzare DSpace ed Arc, ha introdotto nuove funzionalità ed ha rimosso il limite dei due livelli. Per ulteriori informazioni sull'iniziativa italiana si veda [29].

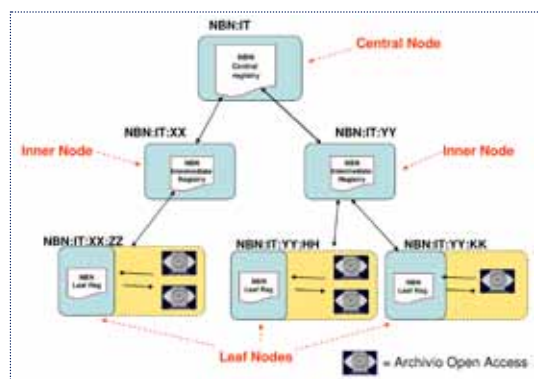


Fig. 1 Architettura gerarchica-distribuita

## 3. Il software jNBN

Nella presente sezione si forniscono alcune informazioni sul software jNBN, rilasciato sotto licenza EUPL<sup>2</sup>.

### 3.1 Componenti architetturali

Il sistema prevede un'organizzazione ad albero dei nodi che gestiscono i diversi sotto-domini. Tali nodi possono essere di tre tipi:

- *Central node*: è il nodo radice della gerar-

<sup>2</sup> La European Union Public License è un tipo di licenza open source definito dall'Unione Europea e compatibile con la General Public License.

chia e corrisponde al dominio di più alto livello. Interroga periodicamente i nodi di secondo livello e conserva le associazioni tra tutti gli NBN generati all'interno dell'infrastruttura e le corrispondenti risorse digitali. Ha il compito di verificare l'unicità e la conformità dei Persistent Identifier alla policy generale. I nodi di secondo livello si registrano presso il nodo centrale.

- *Inner Node*: sono nodi intermedi che non si trovano né al vertice della gerarchia né alla sua base. Gestiscono sotto-domini che aggregano, a livello nazionale, le risorse digitali relative a specifici macro-ambiti (es. Università e Ricerca, Beni Culturali, Multimedia, etc.). Non raccolgono quindi metadati direttamente dai *repository* ma soltanto da nodi di livello immediatamente inferiore. Limitatamente ai rispettivi sotto-domini, effettuano controlli di unicità ed aderenza alle policy locali. Queste ultime includono tutti i vincoli definiti per il dominio di livello immediatamente superiore e ne possono definire di ulteriori.
- *Leaf node*: sono responsabili dei domini di più basso livello associati alle organizzazioni che gestiscono le digital libraries. Sono gli unici nodi che eseguono l'*harvesting* dai *repository* sotto il loro controllo ed assegnano identificativi NBN.

Tutti i tipi di nodo presentano un'interfaccia web e sono in grado di risolvere, direttamente o indirettamente, identificativi appartenenti a qualsiasi sotto-dominio. In particolare, se l'identificativo appartiene al sotto-dominio associato al nodo contattato, questo effettua una ricerca nelle proprie tabelle e fornisce la risposta. In caso contrario, effettua il *parsing* dell'identificativo, individua il *leaf node* competente e, se ne conosce l'indirizzo, gli inoltra la richiesta tramite un particolare web service. Se il *leaf node* interrogato è attivo, fornisce la risposta, che viene poi rigirata all'utente. In caso contrario, il nodo contattato dall'utente risale la gerarchia fino a che non individua un

nodo attivo in grado di rispondere. Al momento attuale sono gli amministratori dei singoli nodi a gestire manualmente le informazioni relative agli altri nodi della rete.

### 3.2 Tecnologie e strumenti

Il software jNBN è basato sulla piattaforma Java Enterprise Edition. Per la sua realizzazione è stato utilizzato il framework per lo sviluppo rapido di applicazioni denominato JADA. Tale framework, realizzato completamente dal Consiglio Nazionale delle Ricerche, permette di generare automaticamente i principali elementi architetturali di un'applicazione Java Enterprise sulla base di configurazioni impostate dallo sviluppatore. Queste configurazioni possono essere specificate o all'interno del codice sorgente (sotto forma di particolari annotazioni) o in appositi file XML.

Per la parte di presentazione è stato utilizzato zKoss, un framework che semplifica l'utilizzo delle primitive messe a disposizione da Ajax, consentendo di creare rapidamente interfacce web avanzate. L'*harvesting* dei metadati viene eseguito mediante il protocollo OAI-PMH.

Il database di default è Postgres, ma è possibile utilizzare anche altri DBMS. Il pacchetto di installazione automatica include un application server JBoss, ma nulla vieta di effettuare il *deployment* dell'applicazione su un qualunque *application server* JEE-compliant. L'infrastruttura di *Software Configuration and Change Management* utilizzata per gestire il processo di sviluppo include un sotto-sistema di *versioning* basato su Subversion ed un sotto-sistema di *change tracking* basato su Bugzilla.

### 3.3 Caratteristiche funzionali

Allo stato attuale jNBN prevede quattro profili utente:

- *amministratore*: svolge compiti di gestione applicativa quali la creazione delle utenze, la pianificazione delle operazioni di *harvesting* e l'inserimento di informazioni relative ad altri nodi;



- *operatore*: gestisce le operazioni di creazione/assegnazione degli identificativi persistenti;
- *utente*: interroga il registro fornendo come input nomi NBN e, ottenendo dal sistema il riferimento alle corrispondenti risorse digitali o ai relativi metadati, accede ad una vista categorizzata delle risorse digitali.

Le funzionalità implementate sono le seguenti:

- *recupero di contenuti e metadati*: fornendo come input un identificativo NBN, l'utente può scegliere di ottenere o la corrispondente risorsa digitale o i metadati ad essa associati;
- *risoluzione di tutti gli identificativi da parte di qualunque nodo della rete*: come specificato precedentemente, ogni nodo della rete è in grado di risolvere qualunque identificativo NBN, anche quelli non appartenenti al proprio sotto-dominio di competenza;
- *pianificazione delle operazioni di harvesting*: l'amministratore del nodo può impostare un numero arbitrario di operazioni di *harvesting* temporizzate per recuperare metadati ad intervalli regolari o dalle digital library (nel caso di *leaf node*) o dai nodi di livello inferiore (nel caso di *inner node* e *central node*);
- *assegnazione dei nomi NBN*: gli operatori dei *leaf node* verificano periodicamente la presenza di nuove risorse digitali e gestiscono l'assegnazione degli identificativi;
- *controllo dei duplicati*: ciascun nodo verifica che all'interno del proprio dominio di competenza non esistano risorse identiche con diversi identificativi; allo stato attuale

ciò viene fatto mediante il confronto di codici *hash* MD5 calcolati a partire dal contenuto delle risorse.

Gli schemi di metadati supportati al momento sono Dublin Core (eventualmente codificato tramite Mets) ed MPEG-21 Digital Item Declaration Language (DIDL).

#### 4. Il testbed italiano: struttura e primi risultati

Al fine di testare le caratteristiche del software jNBN, è stato predisposto un testbed nazionale, all'interno del quale la Biblioteca Nazionale Centrale di Firenze gestisce il nodo centrale (dominio nazionale NBN:IT). Il CNR gestisce un nodo di secondo livello, relativo al settore Università e Ricerca (NBN:IT:UR), ed il proprio *leaf node* (NBN:IT:CNR). La Fondazione Rinascimento Digitale ha il proprio nodo foglia (NBN:IT:FRD), così come l'Università di Milano (NBN:IT:UR:UNIMI).

I primi test hanno mostrato la validità dell'approccio distribuito, che garantisce una maggiore scalabilità e ridondanza rispetto ad un approccio centralizzato. Dall'altro lato, si è evidenziato che i controlli di unicità dei nomi possono essere effettuati tenendo presente le seguenti avvertenze/limitazioni:

- i codici *hash* devono essere calcolati a livello di digital library ed inclusi nel set di metadati "harvestabili", onde evitare il download delle risorse da parte dei *leaf node*;
- i codici MD5 permettono di individuare solo duplicati esatti; risorse che differiscono anche per un solo bit vengono considerate a tutti gli effetti diverse.

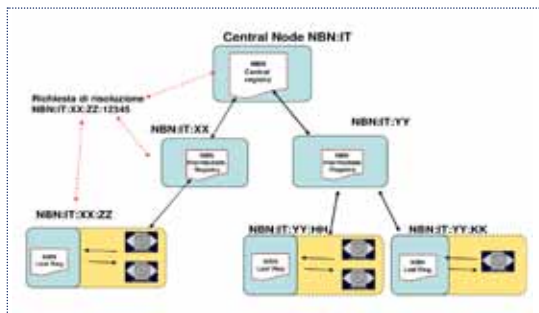


Fig. 2 Invio di richieste di risoluzione a qualunque nodo

#### 5. Sviluppi futuri

Il Consiglio Nazionale delle Ricerche e la Fondazione Rinascimento Digitale, oltre a partecipare alle attività per la sperimentazione di un'infrastruttura nazionale NBN, sono coinvolti con ruoli di rilievo all'interno del progetto PersID, finanziato dal consorzio nord-europeo "Knowledge Exchange" e dalla olandese.

dese “SURF foundation”. Tale progetto ha come scopo quello di realizzare un *global resolver* NBN di livello europeo. In tale contesto il software jNBN è stato considerato un valido punto di partenza per lo sviluppo del sopra citato *global resolver*.

A livello italiano si sta lavorando per allargare il network di sperimentazione. Contatti sono stati intrapresi con le biblioteche nazionali di vari paesi europei per avviare attività congiunte di sperimentazione (ad esempio, la Repubblica Ceca).

Ulteriori possibili evoluzioni del software riguardano l'implementazione di nuove funzionalità quali la possibilità di estrarre metrichi sulla produzione di risorse digitali, sia a livello di sotto-dominio che a livello globale, una migliore capacità di individuare i duplicati e l'interoperabilità con altri standard di identificazione, quale il DOI.

### Riferimenti bibliografici

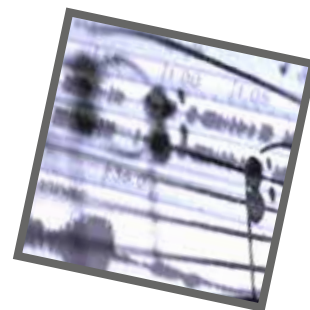
- [1] European Digital Library  
<http://www.theeuropeanlibrary.org/portal/organisation/cooperation/archive/edlproject/>
- [2] Europeana, <http://www.europeana.eu/>
- [3] IETF RFC 3188 Using National Bibliography Numbers as Uniform Resource Names  
<http://tools.ietf.org/html/rfc3188>
- [4] IETF RFC 2141 URN Syntax  
<http://tools.ietf.org/html/rfc2141>
- [5] C. Lagoze and H. V. de Sompel, The Open Archives Initiative Protocol for Metadata Harvesting, version 2.0. Technical report, Open Archives Initiative, 2002. <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [6] Dublin Core Metadata Initiative, Dublin Core Metadata Element Set, Version 1.1  
<http://dublincore.org/documents/dces/>
- [7] MPEG-21, Information Technology, Multimedia Framework, Part 2: Digital Item Declaration, ISO/IEC 21000-2:2003, March 2003
- [8] METS, <http://www.loc.gov/standards/mets/>
- [9] H. Van de Sompel et al. Resource Harvesting within the OAI-PMH Framework D-Lib Magazine December 2004 Volume 10 Number 12 ISSN 1082-9873
- [10] J. Kunze, The ARK Persistent Identifier Scheme, Internet Draft, 2007  
<http://tools.ietf.org/html/draft-kunze-ark-14>
- [11] N. Paskin, Digital Object Identifiers, Inf. Serv. Use, 22(2-3):97–112, 2002  
<http://www.doi.org>
- [12] S. X. Sun. Internationalization of the Handle System. A persistent Global Name Service, 1998  
<http://citeseer.ist.psu.edu/sun98internationalization.html>. [www.handle.net](http://www.handle.net)
- [13] Persistent URL, <http://purl.oclc.org>
- [14] Library of Congress Control Number  
<http://www.loc.gov/marc/lccn.html>
- [15] D. Giaretta. The CASPAR Approach to Digital Preservation. The International Journal of Digital Curation Issue 1, Volume 2, 2007
- [16] S. Andersson, P. Hansson, U. Klosa, E. Muller, E. Siira. Using XML for Long-term Preservation: Experiences from the DiVA Project
- [17] E. Bermes. International Persistent Identifiers for Digital Resources: The experience of the National Library of France. Preservation News, Vol 40 December 2006, pp 23-26  
<http://www.ifla.org/VI/4/news/ipnn40.pdf>
- [18] K. Schroeder. Persistent Identification for the Permanent Referencing of Digital Resources - The Activities of the EPICUR Project Enhanced Uniform Resource Name URN Management at Die Deutsche Bibliothek. The Serials Librarian, 49:75–87(13), 5 January 2006
- [19] CENL Task Force on Persistent Identifiers, Report 2007, [http://www.nlib.ee/cenl/docs/CENL\\_Taskforce\\_PI\\_Report\\_2006.pdf](http://www.nlib.ee/cenl/docs/CENL_Taskforce_PI_Report_2006.pdf)

- [20] K. Sollins. Architectural Principles of Uniform Resource Name Resolution (RFC 2276) <http://www.ietf.org/rfc/rfc2276.txt>
- [21] L. Masinter, K. Sollins. Functional Requirements for Uniform Resource Names (RFC 1737) <http://www.ietf.org/rfc/rfc1737.txt>
- [22] E. Bellini, M. Lunghi, E. Damiani, C. Fugazza. Semantics-aware Resolution of Multi-part Persistent Identifiers, WCKS 2008 conference
- [23] E. Bellini, C. Cirinna, M. Lunghi, E. Damiani, C. Fugazza. Persistent Identifiers distributed system for cultural heritage digital objects, IPRES2008 conference
- [24] H.-W. Hilse, J. Kothe. Implementing Persistent Identifiers: overview of concepts, guidelines and recommendations, 2006, ix+57 pp. 90-6984-508-3 <http://www.knaw.nl/ecpa/publ/pdf/2732.pdf>
- [25] DCC Workshop on Persistent Identifiers, 30 June – 1 July 2005 Wolfson Medical Building, University of Glasgow <http://www.dcc.ac.uk/events/pi-2005/>
- [26] ERPANET workshop Persistent Identifiers, Thursday 17th - Friday 18th June 2004-University College Cork, Cork, Ireland <http://www.erpanet.org/events/2004/cork/index.php>
- [27] Handle System, <http://www.handle.net/>
- [28] Wikipedia Handle page [http://en.wikipedia.org/wiki/Handle\\_System](http://en.wikipedia.org/wiki/Handle_System)
- [29] E. Bellini, C. Cirinnà, M. Lancia, M. Lunghi, R. Puccinelli, M. Saccone, B. Sebastiani, M. Spasiano. Persistent Identifier Distributed System for Digital Libraries, IFLA, 2009, Milano

# Tecnologie basate su XML per la fruizione avanzata dei contenuti musicali

Adriano Baratè, Luca A. Ludovico, Davide A. Mauro

Laboratorio di Informatica Musicale (LIM) - Dipartimento di Informatica e Comunicazione (DICO) - Università degli Studi di Milano



**Abstract.** In questo articolo vengono introdotte le principali caratteristiche di IEEE 1599-2008, uno standard internazionale per la rappresentazione dell'informazione relativa a documenti musicali in XML. Il principale vantaggio offerto dal formato consiste nella possibilità di codificare all'interno di un unico file XML l'informazione eterogenea (simbolica, grafica, audio, video, strutturale) relativa al brano musicale in oggetto. Il presente lavoro mostra inoltre alcune applicazioni notevoli in ambito didattico e di intrattenimento.

## 1. Il formato IEEE 1599-2008

Gli ambiti comunicativi coinvolti nella rappresentazione di un brano musicale possono risultare assai più vasti di quanto emerga da un'analisi superficiale.

I principali aspetti cui un musicista si soffermerebbe nel descrivere una partitura sono quelli cosiddetti simbolici, ossia riconducibili alla disposizione di simboli sul pentagramma e finalizzati all'esecuzione. Viceversa, la descrizione di maggior interesse per un fruitore è legata al segnale audio e di conseguenza all'ascolto.

Tra i due estremi citati si colloca un ventaglio di ulteriori descrizioni eterogenee per un brano musicale: dai metadati catalografici che lo contraddistinguono alla struttura intrinseca del pezzo, dalle informazioni di natura grafica e tipografica legate alle differenti edizioni cartacee ai materiali iconografici collegabili al brano stesso o a una sua esecuzione (bozzetti, figurini, foto di scena, ecc.).

Ciascuno degli aspetti citati trova una controparte in ambito informatico in uno o più formati, aperti o proprietari, standard o legati alle singole applicazioni. Il risultato, dal punto di vista della rappresentazione musicale, è la disponibilità di un vasto insieme di informazioni che spesso sono scorrelate tra loro, codificate in formati eterogenei e dunque dif-

ficilmente fruibili come un *unicum*.

Un'efficace risposta alla problematica di descrivere la musica nel suo complesso, all'interno di un unico documento e beneficiando di tutti i vantaggi propri dell'XML, è data dal formato IEEE 1599-2008. Si tratta di uno standard internazionale recentemente approvato dall'Institute of Electrical and Electronics Engineers (IEEE) il cui obiettivo è fornire una descrizione in XML di singoli brani musicali, prendendo in considerazione tutte le forme di descrizione collegabili al brano stesso. Tale risultato viene conseguito organizzando i dati e i metadati collegati in sei livelli (*layer*):

- *General*, che contiene i metadati relativi al brano in oggetto, tra cui le informazioni catalografiche su titolo dell'opera, autori e genere;
- *Logic*, vero nucleo del formato, destinato alla descrizione simbolica dei contenuti musicali;
- *Structural*, che identifica gli oggetti musicali su cui il brano è costruito e permette di evidenziarne i mutui rapporti;
- *Notational*, che contiene le differenti rappresentazioni grafiche della partitura, ad esempio riferibili a diverse edizioni o trascrizioni;
- *Performance*, che è dedicato ai formati per la generazione di esecuzioni sintetiche da parte

dell'elaboratore;  
 - *Audio*, che consente di legare al brano in oggetto le esecuzioni audio/video della partitura.  
 Nel formato IEEE 1599-2008, tali livelli corrispondono

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ieee1599 SYSTEM "http://www.mx.dico.
unimi.it/ieee1599.dtd">
<ieee1599>
  <general>...</general>
  <logic>...</logic>
  <structural>...</structural>
  <notational>...</notational>
  <performance>...</performance>
  <audio>...</audio>
</ieee1599>
```

Fig. 1 La struttura di un documento IEEE 1599-2008

tecnicamente a sei sotto-elementi dell'elemento radice, come mostrato in figura 1.

Nel creare il formato, una particolare cura è stata posta non solo nel consentire descrizioni qualitativamente eterogenee dei materiali, ma anche nel supportare una molteplicità di materiali dello stesso tipo. Ad esempio, all'interno del livello *Notational* è possibile codificare l'informazione relativa a differenti versioni, manoscritte e a stampa, dello stesso brano; analogamente, nel livello *Audio* è possibile inserire più esecuzioni del brano in oggetto.

Un documento IEEE 1599-2008 sarà tanto più interessante dal punto di vista della rappresentazione e della fruizione quanto più esso risulterà ricco di contenuti. E tale ricchezza può esprimersi tanto nella compilazione di una molteplicità di livelli, quanto nell'inserimento di una molteplicità di oggetti digitali in ciascun livello.

Un altro aspetto di fondamentale importanza nell'approccio descritto è il pieno supporto ai formati già comunemente in uso per i contenuti multimediali. Infatti in XML vengono esplicitamente descritti solo i metadati e le informazioni simboliche sulla partitura, mentre per l'audio, il video e la grafica si codificano gli opportuni riferimenti a file esterni. In tal modo si beneficia del corpus di oggetti digitali già disponibile (ad esempio, dei molti file MP3 relativi a una data canzone) e delle peculiarità di ciascun formato nel fornire una descrizione opportuna per i contenuti multimediali (ad esempio, delle caratteristiche del formato JPEG o TIFF per la grafica).

Il livello *Logic*, che come già detto rappresenta il nucleo del formato, presenta una struttura dati chiamata *spine* il cui obiettivo è elencare tutti gli eventi musicali contenuti all'inter-

no del brano, o più propriamente elencare gli eventi di interesse per chi effettua la codifica. Ad esempio, si può pensare a tutte le note e le pause sul pentagramma, ma su questo aspetto il formato non è restrittivo. Lo *spine*, costituito essenzialmente da un elenco di eventi, ha una duplice funzione: ordina mutuamente gli eventi stessi e li etichetta in modo univoco. Ogni altro livello conterrà descrizioni o riferimenti agli eventi basandosi sull'identificativo univoco dello *spine*. Per quanto concerne i dati multimediali contenuti nei file esterni, il formato IEEE 1599-2008 non li descrive *ex novo* bensì attua un'opportuna mappatura degli eventi dello *spine* (tramite la loro marcatura univoca) su coordinate spaziali o temporali riferite al formato adottato. Ad esempio, si consideri una nota identificata nella struttura dati dall'etichetta eventoN; per collocarla nelle diverse tracce presenti nel livello *Audio* sarà sufficiente esprimere in XML la relazione tra eventoN e il millisecondo o il frame relativo negli audio. Analogamente, per mostrare l'occorrenza della nota nei file grafici relativi alla partitura autografa o a una versione a stampa, si potrà legare nel livello *Notational* l'etichetta eventoN alle coordinate dei punti che ne delimitano l'area occupata.

Il meccanismo che demanda a file esterni la descrizione multimediale degli eventi e che si basa su riferimenti allo *spine* è molto efficace ed efficiente, in quanto consente di aggiungere un numero arbitrario di nuovi documenti con costi lineari rispetto alla sincronizzazione con quanto già presente. In altri termi-

ni, aggiungere una traccia audio o le scansioni di un'ulteriore partitura non comporta la creazione di riferimenti con tutti gli oggetti già mappati, ma solo la determinazione dei punti o degli istanti in cui occorrono gli eventi musicali listati nello *spine* all'interno del nuovo file.

Per ulteriori approfondimenti, le caratteristiche del formato sono illustrate in maggior dettaglio in [1].

## 2. Applicazioni di IEEE 1599-2008

Un formato potenzialmente ricco quanto IEEE 1599-2008 consente un gran numero di applicazioni nei campi più svariati dell'informatica musicale. A puro titolo di esempio, si può citare la modellazione di brani musicali secondo metriche matematiche [2], l'interazione in tempo reale con le strutture musicali [3], la sincronizzazione automatica dei contenuti multimediali [4].

Innanzitutto, il formato può rivestire un ruolo importante nella didattica musicale, grazie allo sviluppo di strumenti che fungono sia da guida in contesti di fruizione musicale, sia come ausilio per l'*ear training*. Da questo punto di vista, l'eterogeneità dei materiali descritti nei *layer* del formato IEEE 1599-2008 consente di sviluppare applicazioni multimediali in cui gli aspetti dell'informazione musicale possono essere fruiti a vari livelli. Un primo utilizzo di base permette agli utenti di concentrarsi sulla sincronizzazione audio/grafica per apprendere la lettura di musica mentre si ascolta una sua esecuzione. Le possibilità più avanzate consistono invece nel cambiare in tempo reale – quindi senza bloccare l'esecuzione – i vari file audio o video per confrontare esecuzioni diverse dello stesso brano. Quan-

to detto sui materiali audio può essere esteso anche ai documenti grafici: spesso risulta interessante comparare i manoscritti, da cui si evince il processo compositivo, con le edizioni a stampa, in cui la grafica è spesso più leggibile e meno tormentata. Oltre alla possibilità di scegliere in real time la traccia audio o la rappresentazione grafica corrente, un software basato su IEEE 1599-2008 consente altri gradi di interazione; ad esempio cliccando sulla partitura di un brano la traccia audio viene risincronizzata alla nota selezionata. È importante sottolineare che il formato è compatibile anche con rappresentazioni grafiche non tradizionali, usate per esempio nella musica elettronica.

Quanto detto finora trova applicazione nei numerosi software basati su tale tecnologia e sviluppati per mostre e convegni. Ad esempio, in occasione della conferenza "Il nuovo in musica - Estetiche tecnologie linguaggi" (Trento, 18-20 gennaio 2008), un applicativo IEEE 1599-2008 (*fig.2*) è servito come supporto alla dimostrazione di una tesi musicologica di Rossana Dalmonte [5]. Esso mostra la partitura di un duetto tratto da *Tosca* di Giacomo Puccini, consentendone la visualizzazione sincronizzata con l'ascolto, e ne evidenzia con chiarezza l'andamento melodico a grandi arcate anche quando queste risultano distribuite tra le due voci principali o frammentate nella tessitura orchestrale.



Fig. 2 Dimostrativo su *Tosca* di Giacomo Puccini



Un'applicazione a fini didattici della tecnologia IEEE 1599-2008 è pure presente presso il Museo degli Strumenti Musicali del Castello Sforzesco di Milano. In tal caso (fig.3), si tratta di rendere fruibile al pubblico un'analisi musicologica articolata e complessa realizzata da Angela Ida De Benedictis su alcune opere di Bruno Maderna che presentano una matrice comune [6].

Le caratteristiche del formato che trovano applicazione in ambito didattico possono essere validamente sfruttate anche nel campo dell'intrattenimento. Ad esempio, la possibilità di collegare a un'unica entità-brano varie rappresentazioni sincronizzate e sostituibili in tempo reale risponde tanto alle esigenze dell'appassionato di musica colta, desideroso di confrontare esecuzioni ed interpreti, quanto agli interessi del fruitore di musica leggera, cui il formato può offrire varie versioni del brano e numerosi materiali collegati. A questo riguardo, esistono numerosi esempi commissionati da enti quali il Ministero per i Beni e le Attività Culturali, l'Archivio Storico Ricordi e la RSI – Radiotelevisione Svizzera (fig.4).

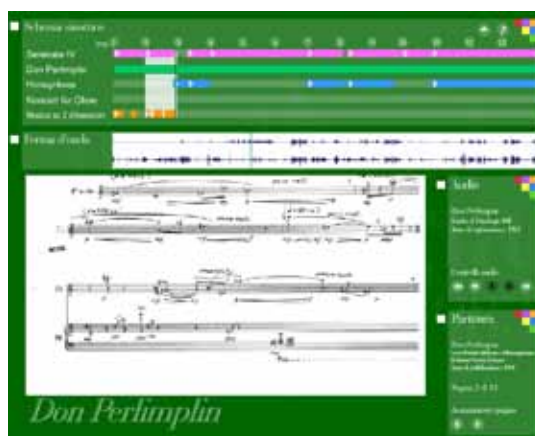


Fig. 3 Dimostrativo su alcune opere di Bruno Maderna

### 3. Conclusioni

In questo articolo sono state presentate le principali caratteristiche dello standard IEEE 1599-2008 orientate all'intrattenimento e alla fruizione di contenuti multimediali.

Come mostrato, gli strumenti software basati su IEEE 1599-2008 permettono di seguire un brano musicale nei suoi molteplici aspetti tra loro complementari: la partitura simbolica, la notazione grafica, gli audio, i materiali correlati e persino le strutture intrinseche.

Uno dei punti di forza consiste nel poter

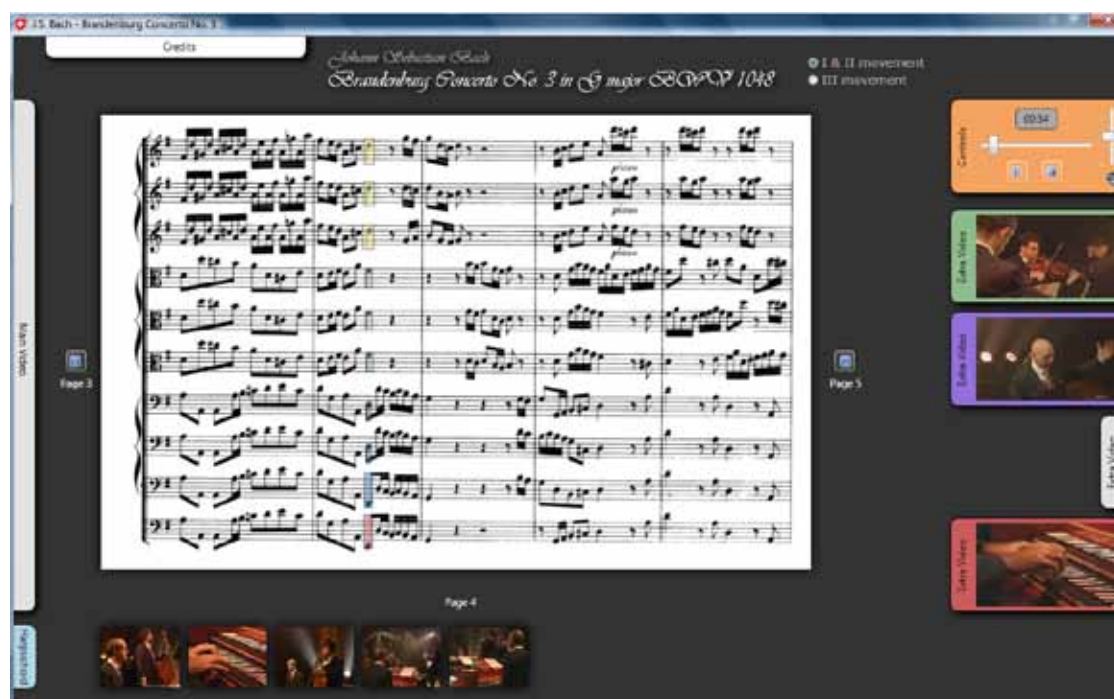


Fig. 4 Dimostrativo sui *Concerti Brandeburghesi* di Johann Sebastian Bach

presentare in maniera intuitiva i contenuti musicali anche a un fruitore senza specifica formazione. Inoltre, la profonda integrazione tra oggetti multimediali eterogenei rafforza la comprensione delle caratteristiche musicali, delle strutture e delle interpretazioni.

In conclusione, lo standard IEEE 1599-2008 apre nuove frontiere per la fruizione e l'interazione con i contenuti musicali, in particolare nei campi dell'editoria elettronica e dell'intrattenimento. Tra le applicazioni innovative di tale formato, si prefigura l'integrazione all'interno di prodotti multimediali quali lettori DVD, console orientate all'intrattenimento e dispositivi portatili di ultima generazione (smartphone, PDA, lettori multimediali, ecc.).

La pubblicazione di documenti conformi allo standard IEEE 1599-2008 offre una descrizione più ricca dei contenuti musicali e la rende disponibile a un vasto pubblico.

Ulteriori informazioni sul formato e sui progetti relativi sono disponibili nel sito web: <http://www.mx.dico.unimi.it>.

## Riferimenti bibliografici

[1] L.A. Ludovico, "Key Concepts of the IEEE 1599 Standard", Proceedings of the IEEE CS Conference The Use of Symbols To Represent Music And Multimedia Objects, IEEE CS, Lugano, Confederazione elvetica, 2008

[2] A. Pinto, "Multi-model Music Content Description and Retrieval Using IEEE 1599 XML Standard", Journal of Multimedia (JMM), Special Issue on The New Standard IEEE 1599 - Interacting with Music Contents by XML Symbols, Volume 4, Issue 1, Febbraio 2009

[3] A. Baratè, "Real-time Interaction with Music Structures in IEEE 1599", Journal of Multimedia (JMM), Special Issue on The New Standard IEEE 1599 - Interacting with Music Contents by XML Symbols, Volume 4, Issue 1, Febbraio 2009

[4] A. D'Aguanno, G. Vercellesi, "Automatic synchronisation between audio and score musical description layers", Proceedings of the 2nd international conference on Semantics And digital Media Technologies (SAMT 2007), Genova, Dicembre 2007

[5] R. Dalmonte, "Analisi melodica e tecnologia. Un esempio da Tosca di Puccini", Atti del Convegno Il nuovo in musica - Estetiche tecnologie linguaggi, LIM, Trento, Gennaio 2008

[6] A.I. De Benedictis, "Materiali parziali o strutture fungibili? Alcune riflessioni su Honeymèves, Don Perlimplin e Serenata IV", Relazione alle Giornate di Studio su Bruno Maderna, Università degli Studi di Torino, Torino, Novembre 2007

# Infrastruttura telematica per la gestione informatizzata del blocco operatorio

Alessandro Pepino<sup>1</sup>, Tullio Cusano<sup>2</sup>

<sup>1</sup> DIBET Università degli Studi di Napoli Federico II,

<sup>2</sup> Azienda Ospedaliera di Rilievo Nazionale e di Alta specializzazione, Vincenzo Monaldi di Napoli



**Abstract.** L'articolo illustra le caratteristiche del nuovo modello organizzativo della Azienda Ospedaliera Monaldi e delle nuove soluzioni informatiche/telematiche adottate. In particolare, il modello organizzativo ha previsto da un lato una stretta sinergia logistica ed organizzativa tra le sale operatorie e la Rianimazione/Terapia intensiva a garanzia della sicurezza del paziente, dall'altro ha richiesto una completa revisione degli assetti organizzativi e quindi delle soluzioni informatiche. Dal punto di vista telematico, il complesso operatorio è cablato in modo capillare al fine di garantire la massima flessibilità operativa e prevede una serie di connessioni analogiche audio/video facenti capo ad una sala regia centralizzata che consente di predisporre complessi eventi multimediali dal vivo su più sale operatorie in contemporanea.

## 1. Introduzione

Negli ultimi anni si è ormai consolidata la consapevolezza che nelle cosiddette “aree critiche” (Sale operatorie e Terapie intensive) i processi devono essere oggetto di un attento controllo ai fini da un lato della razionalizzazione dei costi e dall'altro della riduzione dei rischi connessi alle attività mediche e infermieristiche. Peraltro le aree critiche sono anche fonte inesauribile di conoscenza ed esperienza pratica che, per essere adeguatamente sfruttate a fini formativi e di aggiornamento scientifico, richiedono appropriate infrastrutture tecnologiche per la cattura, gestione e distribuzione dei flussi audio e video relativi alle attività chirurgiche. La ricerca su “pube” con le seguenti keyword “*operating theatre information system*” ha dato circa 479 risultati di cui numerosi relativi a contributi strettamente inerenti ai temi toccati in questo lavoro e pubblicati in questi ultimi anni.

Nel complesso, la maggior parte dei lavori scientifici così selezionati, di cui solo una parte è citata in bibliografia, concordano nel ritenere strategici i seguenti obiettivi:

- ottimizzazione dei tempi di occupazione sale;

- tracciatura e controllo dei consumi;
- riduzione del rischio clinico;
- sfruttamento della conoscenza ed esperienza prodotta in sala operatoria.

Nel complesso tale bibliografia riporta quindi diverse esperienze di uso dei sistemi informativi finalizzati agli obiettivi indicati che, al di là della specifiche soluzioni software adottate, (rilevante solo entro certi limiti) hanno diversi elementi chiave in comune:

- impiego di touch screen in sala operatoria;
- riconoscimento del paziente tramite barcode o RFID;
- acquisizione online dei parametri vitali o addirittura dei segnali elettrofisiologici;
- adozione di sistemi di videoconferenza nelle sale operatorie.

Il presente lavoro intende quindi riportare, a titolo di ulteriore esempio svolto, il complesso sistema in funzione presso l'A.O. Monaldi, frutto di un'esperienza sul campo maturata in circa 10 anni di attività; quest'ultima esperienza risulta particolarmente significativa in quanto riassume in un unico sistema integrato e gestito in modo coordinato, diversi sistemi tecnologici complessi, il tutto in un contesto ambientale particolarmente complesso e diffi-

cile quale quello di una grande Azienda Ospedaliera pubblica del sud.

## 2. Architettura del sistema

Sebbene progettata negli anni tra il 1998 e il 2000, la Piastra operatoria della A.O. Monaldi conteneva, già al momento della sua inaugurazione nel 2004, una serie di elementi propedeutici alle soluzioni tecnologiche che sono state realizzate negli anni a venire e che non sarebbero state realizzabili in assenza di una adeguata infrastruttura.

Ubicata nella zona centrale del presidio ospedaliero, la Piastra è composta da 10 sale operatorie ed è dotata anche di un'unità di terapia intensiva cardio-respiratoria (TICR) ed un'unità di terapia intensiva post-operatoria (TIPO). Nelle nuove sale operatorie vengono eseguiti attualmente gli interventi di urologia, ortopedia, chirurgia generale, chirurgia vascolare e chirurgia toracica, otorinolaringoiatra e oculistica.

re diversi flussi di dati tra cui: segnali elettrofisiologici, video-audio real time ad alta definizione, dati e immagini da e verso il sistema informativo.

Naturalmente occorre che i sistemi legati alla loro produzione, gestione e fruizione rispettino requisiti stringenti di affidabilità, sicurezza/riservatezza e semplicità di uso più che in qualsiasi altro comparto di un sistema telematico ospedaliero.

L'infrastruttura telematica pertanto prevede un cablaggio composto da tre diverse tipologie di reti:

- Rete di monitoraggio
- Rete dati
- Rete audio-video

### 2.1 Sistema integrato di monitoraggio

La rete di monitoraggio è una semplice fast ethernet composta da tre subnet fisiche collegate tra di loro e separate sia a livello fisico che logico dalla rete aziendale. Essa consente il collegamento di tre serie di monitor multipa-

rametrici (Sale operatorie, Terapia Intensiva CardioRespiratoria: TI-CR e Terapia Intensiva post Operatoria: TIPO ) ad altrettante postazioni computerizzate denominate "centrali", dedicate alla sorveglianza dei parametri elettrofisiologici. I dati provenienti dalle stazioni di monitoraggio vengono memorizzati e gestiti su un Data Server su cui è disponibile un apposito servizio http che rende possibile, da un qualun-

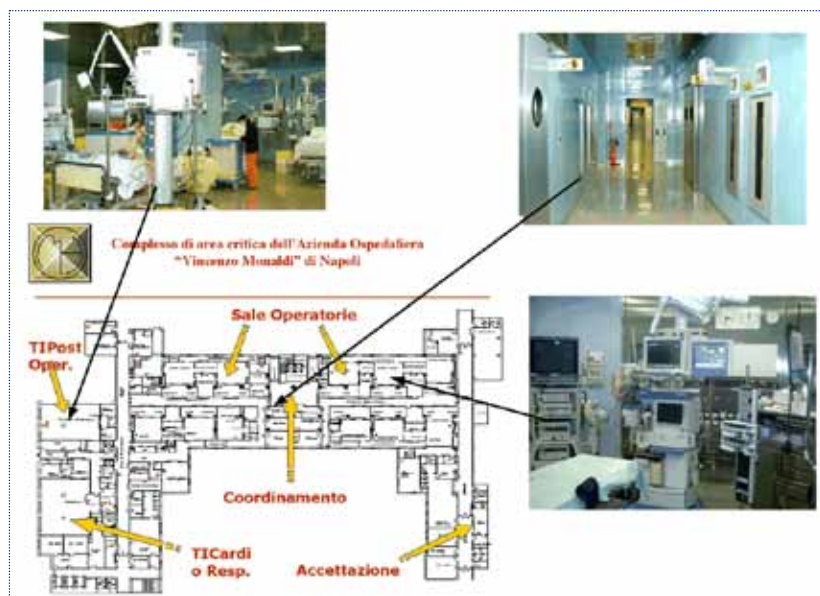


Fig. 1 Area critica dell'A.O. V. Monaldi

La Piastra è un comparto operativo di elevata complessità, ove le esigenze di trasmissione e gestione dei dati sono fortemente diversificate e complesse. Per venire incontro alle esigenze citate in premessa, occorre infatti gesti-

re diversi flussi di dati tra cui: segnali elettrofisiologici, video-audio real time ad alta definizione, dati e immagini da e verso il sistema informativo.

storia degli eventi relativi al singolo paziente. Grazie poi ad un apposito sistema di integrazione con i database aziendali, dai monitor multiparametrici al posto letto, è possibile anche accedere ai dati relativi agli esami del laboratorio. Particolare attenzione è stata infine prestata alla sicurezza del riconoscimento del paziente al fine della riduzione del rischio clinico: l'identificazione del paziente nella rete di monitoraggio viene infatti effettuata attraverso il riconoscimento del codice nosografico del paziente, che viene stampato sotto forma di barcode su di un apposito braccialetto indossato dal paziente sin dall'atto del ricovero.

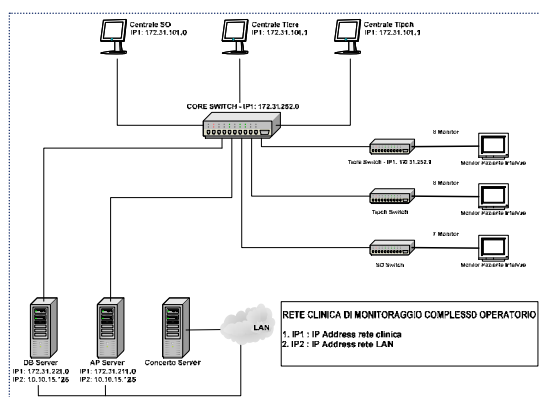


Fig. 2 Layout del sistema di monitoraggio di parametri vitali

## 2.2 Rete audio e video - videoconferenza

Le sale operatorie della A.O. Monaldi, nelle diverse specialità coinvolte, sono una inesauribile fonte di conoscenze e di esperienza: il supporto audio-video è quindi un tramite indispensabile per la formazione delle professionalità mediche e la comunicazione scientifica, sia all'interno che all'esterno dell'ospedale. L'uso oramai sempre più esteso dei sistemi di video-laparoscopia permette/richiede una gestione estesa di segnali video provenienti anche da più sorgenti contemporaneamente, con relativi problemi di regia, trasmissione e fruizione. L'alta qualità poi ottenibile dalle moderne telecamere, richiede strumenti di *encoding* video sempre più sofisticati ed una banda trasmissiva sempre maggiore.

Per tali motivi, il progetto della piastra operatoria della A.O. Monaldi ha previsto una completa rete di collegamenti audio/video in tutte le sale operatorie e tra le medesime e una sala regia centralizzata: in ogni sala operatoria è stato previsto un set di connessioni audio e video, sia sulla lampada scialitica che a pannello sulla parete attrezzata, al fine di poter collegare telecamere esterne e apparati di videoendoscopia.

I collegamenti audio-video fanno capo ad una apposita matrice che consente le necessarie permutazioni per smistare ingressi e uscite audio e video, a cui a sua volta è collegato il mixer audio-video. In uscita al mixer, i segnali live possono essere inviati ad un videoregistratore digitale, e/o un sistema di videoconferenza e/o un sistema di streaming video in funzione delle esigenze dello specifico evento. La A.O. Monaldi, inoltre, partecipa insieme al GARR ad un progetto internazionale di sperimentazione di un innovativo sistema di videoconferenza su IP denominato DVTS che consente elevati standard qualitativi, naturalmente a prezzo di un rilevante impegno di banda.

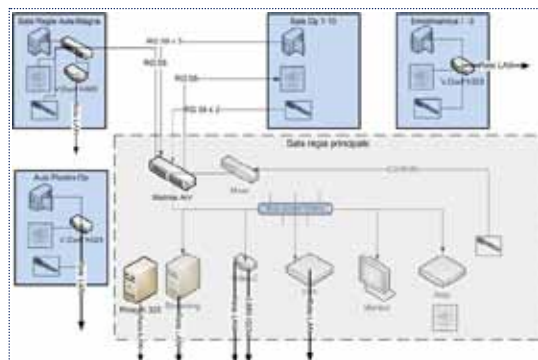


Fig. 3 Layout della rete audio-video

Tramite il sistema di cablaggio strutturato, disponibile presso la A.O. Monaldi, attraverso apposite permutazioni di armadio, è stata quindi predisposta una tratta in fibra dedicata tra la sala regia della piastra operatoria e la connessione GARR presente al CED della A.O. Monaldi da cui il flusso viene poi gestito dalla rete nazionale. La possibilità di gestire



flussi video-audio ad alta qualità su IP apre numerosi scenari applicativi: a tutt'oggi infatti la videoconferenza viene gestita prevalentemente tramite sistemi che adoperano linee ISDN o connessioni via satellite proprio a causa della difficoltà di garantire la qualità del servizio attraverso le normali reti pubbliche. Obiettivo della sperimentazione è quindi quello di costituire attraverso il GARR una piattaforma permanente di condivisione di esperienze live nel settore della chirurgia laparoscopica.

Da un punto di vista operativo, un sistema siffatto da un lato consente una elevata flessibilità d'uso ed ampie potenzialità, dall'altro richiede specifiche competenze sia per la gestione delle apparecchiature di mixaggio ma anche e soprattutto per le attività di ripresa video e di regia.

Solo una sistematica programmazione di eventi ed un uso intensivo del sistema consentono una gestione sostenibile attraverso l'impiego di risorse umane interne. In alternativa, dato il permanere del carattere di saltuarietà degli eventi, è indispensabile fare ricorso a risorse di ditte esterne per le quali il problema è quello di coordinare e gestire le risorse strumentali al fine di ridurre i costi nel rispetto degli standard qualitativi.

### 2.3 Rete Dati – Sistema informativo

Per la gestione dei dati la piastra operatoria dispone di una ulteriore e capillare rete dati in fast ethernet, separata fisicamente da quella già citata del sistema di monitoraggio e ben integrata con la complessa rete telematica della A.O. Monaldi.

L'intero sistema dispone di circa 40 armadi di piano connessi in gigabit su fibra e tre nodi di edificio a loro volta connessi in 10 gb al nodo di campus. La rete dati della piastra operatoria fa quindi capo ad un armadio di piano disposto come buona parte delle attrezzature telematiche della piastra operatoria nella sala regia. Complessivamente sono stati previsti all'interno della piastra operatoria oltre 50 punti rete dislocati soprattutto nelle sale

operatorie e nelle sale di rianimazione e terapia intensiva, ma anche nelle sale studio, nelle sale di coordinamento ed in generale in tutti gli ambienti della piastra operatoria. Sulla rete dati si appoggia il sistema informativo che è costituito da una rete di client distribuiti in tutte le sale operatorie e nei vari punti di accesso e governo dell'area critica, i quali fanno riferimento ad un server dislocato anch'esso all'interno del CED della A.O. Monaldi.

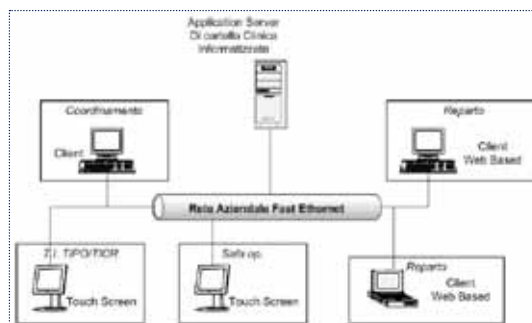


Fig. 4 Layout del sistema informativo

La gestione informatica del blocco operatorio coinvolge un ampio numero di problematiche ed aspetti tecnico-organizzativi. Le principali aree di intervento sono:

- *organizzazione e scheduling delle risorse*: pianificazione degli interventi chirurgici (liste di attesa interne), supervisione delle attività in tempo reale, centrale di gestione della programmazione;
- *controllo di gestione*: collegamento con sistemi informativi aziendali (anagrafe, farmacia, laboratorio, repository centrale), rilevamento dettagliato dei consumi;
- *sicurezza e percorso diagnostico terapeutico*: cartella anestesilogica, acquisizione dati online dalle apparecchiature, rilevamento degli artefatti e validazione dei dati, gestione cartella pre-, intra- e post-operatoria, compilazione Registri operatori.

In ciascuna di queste aree il supporto dei sistemi informativi è fondamentale, dato che esso consente un controllo analitico del processo e quindi da una parte una drastica riduzione dei rischi connessi a errori umani (rischio clinico)



e dall'altra una esatta valutazione dei costi connessi sia in termini di personale che di materiali (presidi e farmaci) in relazione al singolo paziente ed al singolo intervento.

Il sistema informatico prevede da un lato postazioni client basate su di una interfaccia touch screen ottimizzata per l'impiego nelle sale operatorie e in rianimazione, dall'altro numerosi client web fruibili direttamente presso i reparti della A.O. Monaldi. Naturalmente il sistema Touch Screen, da un punto di vista elettrico e in quanto adoperato all'interno dell'Area Paziente, deve essere conforme alle normative di sicurezza per l'uso in diretto contatto col paziente. Come tutti sistemi informativi della A.O. Monaldi, anche quello in uso nella piastra operatoria per la gestione del flusso del paziente utilizza il sistema di riconoscimento paziente basato sul braccialetto con barcode.

Un'ulteriore ed utilissima postazione è costituita da un Carrello Farmaci Informatizzato, denominato STEFI (*Sistema TElematico per la Funzione Infermieristica*) messo a punto da una Ditta di settore nell'ambito di uno specifico programma di ricerca finanziato dalla Regione Campania (POR mis 3.17 progetto Radionet) nel quale il Monaldi è stato sede di sperimentazione. Il sistema, grazie ad una connessione wireless, consente di attivare una serie di funzioni in diretta prossimità del letto del paziente.

Il flusso principale del sistema informatico impiegato all'interno della piastra operatoria è riportato sinteticamente, sotto forma di diagramma UML, nelle figure 6 e 7.

I punti di forza del sistema sono essenzialmente nelle seguenti funzioni:

- *pianificazione degli interventi* che viene effettuata attraverso il client web direttamente nelle U.O. di provenienza del paziente;



Fig. 5 Carrello informatizzato STEFI

- *riconoscimento del paziente* all'ingresso della giusta sala operatoria;
- *tracciatura delle varie fasi operatorie* con registrazione dei tempi di inizio e fine;
- *registrazione del personale* coinvolto nell'intervento;
- *tracciatura e registrazione*

*del consumo* dei presidi e farmaci utilizzati durante l'intervento;

- *stampa del referto operatorio* e pubblicazione in un apposito *repository* conforme agli standard internazionali CDA H17 finalizzato a future integrazioni con la rete regionale.

Al sistema informativo della piastra è connessa poi anche la cartella anestesiologicala. Da questo punto di vista, infatti, l'anestesista è il maggiore produttore e consumatore di dati sia durante che dopo l'intervento. È quindi indispensabile considerare la postazione di anestesia come parte essenziale di ogni sistema di gestione delle sale operatorie. Tutti i tempi operatori, tutti i farmaci, molti materiali di alto costo, le procedure anestesiologicalhe, le complicanze, sono già oggi contenuti nella cartella anestesiologicala



Fig. 6 Use case dello schema organizzativo

cartacea e quindi la sua informatizzazione li rende fruibili per scopi gestionali e di controllo indispensabili nel decorso post-operatorio. La cartella anestesilogica viene utilizzata, naturalmente, sia durante l'intervento che nel post-intervento nelle sale di Terapia Intensiva TIPO e TICR.

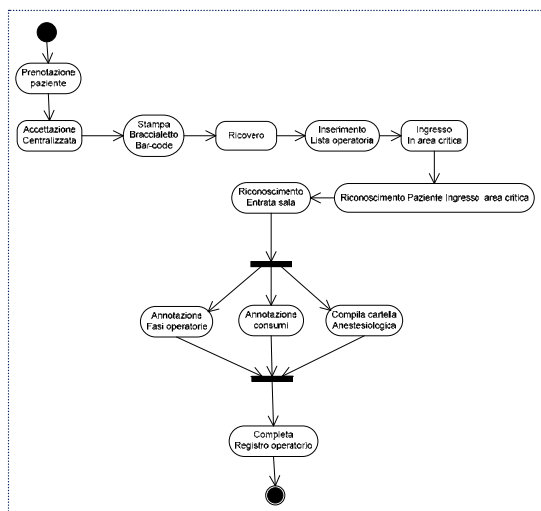


Fig. 7 Activity diagram del sistema informativo

Soprattutto nelle Terapie Intensive, un'utile caratteristica della cartella anestesilogica informatizzata è rappresentata dall'acquisizione online dalle apparecchiature, inclusi monitor, sistemi di anestesia, pompe infusionali, emogasanalizzatore, sistemi per la circolazione extracorporea, analizzatori di gas, gittata cardiaca, monitor della profondità dell'anestesia (BIS, AEP), etc. Tale acquisizione deve essere automatica e continua, ad intervalli sufficientemente brevi (ad esempio ogni 60 secondi).

L'acquisizione automatica dei dati è caratteristica certamente gradita a qualsiasi utilizzatore, sia per il risparmio di tempo rispetto alla copiatura manuale dei dati sulla cartella cartacea, sia per la possibilità di registrare correttamente i dati anche nelle fasi di induzione e risveglio e nelle situazioni di emergenza, quando l'anestesista è ben più impegnato ad occuparsi del paziente piuttosto che della cartella.

### 3. Conclusioni

A valle della realizzazione di un sistema così ricco di funzionalità e potenzialità quale quello realizzato all'interno della A.O. Monaldi, la principale difficoltà risiede nel:

- modesto livello di alfabetizzazione del personale infermieristico;
- resistenza da parte del personale medico verso i nuovi modelli organizzativi.

Nel caso del sistema informativo della Piastra Operatoria, la Direzione Generale ha esibito una forte determinazione che ha consentito di predisporre i necessari punti di controllo al sistema organizzativo, propedeutici all'impiego efficace del sistema informativo. In altri termini il regime di controllo è stato posto come prerequisito per l'accesso alle risorse (farmaci presidi) e il corretto uso del sistema informatico è diventato uno degli obiettivi previsti dalla normativa per la Dirigenza Medica.

Grazie a questo sforzo, organizzativo e politico, è stato possibile innescare e condurre il difficile processo di cambiamento per cui oggi è possibile una efficace programmazione delle attività del comparto operatorio, la registrazione dei tempi relativi alle singole fasi operatorie, il controllo di gestione in tempo reale dei dati relativi al consumo di presidi, farmaci e personale per singolo paziente e per singolo intervento, l'acquisizione online dei parametri vitali, il controllo online dei parametri infusionali.

Questa grande mole di dati, se da un lato consente una dimensione di controllo dei processi impensabile prima della introduzione del sistema informatico e che incide in modo determinante sulla sicurezza del paziente, dall'altro richiede naturalmente un ulteriore sforzo di analisi da parte del management aziendale, al fine di ottenere gli auspicati benefici in termini di razionalizzazione di impiego delle risorse.

### Riferimenti bibliografici

- [1] Dexter F, Marcon E, Aker J, Epstein RH,

Numbers of simultaneous turnovers calculated from anesthesia or operating room information management system data. In *Anesth Analg*. 2009 Sep;109(3):900-5

[2] Meyfroidt G, How to implement information technology in the operating room and the intensive care unit. *Best Pract Res Clin Anaesthesiol*. 2009 Mar;23(1):1-14

[3] Chow VS, Huang W, Puterman ML, Benefits and challenges of health information systems for operations research: an illustrative example to improve surgical scheduling in *Stud Health Technol Inform*. 2009;143:24-9

[4] Park KW, Dickerson C, Can efficient supply management in the operating room save millions?in *Curr Opin Anaesthesiol*. 2009 Apr;22(2):242-8

[5] Torregrossa MV, Trapani S, Cardinale AE, Evaluation of operating theatre for the risk prevention in *Ann Ig*. 2008 May-Jun;20(3):243-9

[6] Nagy PG, Konewko R, Warnock M, Bernstein W, Seagull J, Xiao Y, George I, Park A, Novel, Web-based, information-exploration approach for improving operating room logistics and system processes. In *Surg Innov*. 2008 Mar;15(1):7-16

[7] Russell J, How technology solutions can impact nursing retention. In *Nurs Econ*. 2008 May-Jun;26(3):188-90

[8] Bosman RJ, Impact of computerized information systems on workload in operating room and intensive care unit. In *Best Pract Res Clin Anaesthesiol*. 2009 Mar;23(1):15-26

# Gestione semantica ed adattativa per la fruizione dei Contenuti Culturali nel Progetto LC3<sup>1</sup>

Rocco Aversa<sup>1</sup>, Beniamino Di Martino<sup>1</sup>, Francesco Moscato<sup>2</sup>,  
Salvatore Venticinque<sup>1</sup>, Pasquale Cantiello<sup>1</sup>

<sup>1</sup> Dipartimento di Ingegneria dell'Informazione, Seconda Università di Napoli

<sup>2</sup> Dipartimento di Studi Europei e Mediterranei, Seconda Università di Napoli



**Abstract.** L'articolo presenta le attività della Seconda Università di Napoli nell'ambito del progetto LC3 (Laboratorio Pubblico-Privato per la Comunicazione delle Conoscenze Culturali in Rete), finanziato dal MIUR. Obiettivo di LC3 è quello di realizzare nuove funzionalità per valorizzare la conoscenza e la fruibilità dei patrimoni culturali a livello nazionale e internazionale. Il gruppo di ricerca in Informatica della Seconda Università di Napoli si è dedicato in particolare alle attività rivolte all'elaborazione di linee guida e soluzioni per la creazione e gestione uniforme di basi di conoscenza da utilizzarsi per annotare semanticamente, a vari livelli, le diverse fonti culturali trattate, in particolare: studio e definizione di un modello per la creazione di annotazioni semantiche; derivazione manuale ed automatica di ontologie di dominio; sistemi di annotazione semantica; definizione di tecniche semantiche per l'*Information Retrieval*; progettazione e realizzazione di un servizio di profilazione ad agenti adattabile al particolare dispositivo.

## 1. Introduzione

Ricercare in rete le informazioni rilevanti è spesso un lavoro né piacevole né semplice, dal momento che i contenuti non sono organizzati secondo criteri semantici, cioè che tengano conto del loro reale significato e del contesto al quale appartengono. Una tale organizzazione infatti garantirebbe un recupero efficace e accurato.

Attualmente le informazioni in Internet non hanno un significato. Con ciò si intende dire che il significato non è un attributo dell'informazione che le applicazioni riescono a considerare nelle loro elaborazioni. È dunque importante, al fine di migliorare il reperimento delle informazioni con query mirate, o per rifinire i risultati delle ricerche, o anche per strutturare i documenti, ricorrere al trattamento non solo sintattico, ma anche semantico dei contenuti, sia per i documenti testuali, sia per quelli multimediali.

Un'erogazione adattativa di contenuti multimediali, che renda possibile ottimizzare il livello di qualità percepito dall'utente in funzione delle risorse disponibili sul termina-

le, dei parametri di sessione, del profilo sociale e culturale dell'utente stesso, è auspicabile sia per gestire l'innovazione tecnologica nei dispositivi, sia per dar vita a nuove modalità di fruizione dei contenuti. Essa può essere utilizzata per supportare persone disabili nell'accesso ai servizi o per adattare la modalità di fruizione al contesto in cui ci si trova. Un testo può essere distribuito come uno stream audio mediante un particolare adattatore, quando il terminale non possiede una interfaccia comoda per la lettura, quando l'utente è un non vedente, o semplicemente sta guidando. Inoltre la *context awareness*, ovvero la sensibilità al contesto da parte di un'applicazione, può aiutare gli utenti a filtrare la grande quantità di dati disponibili in Internet, ma che non sono utilizzabili, o semplicemente non sono rilevanti rispetto al particolare contesto di lavoro.

Tale caratteristica avanzata consente ai fornitori di attrarre un maggior numero di clien-

.....  
<sup>1</sup> Questo lavoro è supportato da LC3 - Laboratorio pubblico-privato di ricerca sul tema della Comunicazione delle Conoscenze Culturali - Progetto nazionale finanziato dal Ministero della Ricerca MIUR DM1791

ti offrendo servizi che ottimizzano il livello di soddisfazione dell'utente. Inoltre, il bisogno di rendere i contenuti del *Semantic Web* accessibili agli utenti sta diventando sempre più pressante, dato che l'ammontare delle informazioni immagazzinate nelle basi di conoscenza di tipo ontologico e le annotazioni semantiche basate su ontologie aumentano costantemente. Le ontologie, che rappresentano la tassonomia e le relazioni che intercorrono all'interno di un dato dominio di interesse (tramite classi e proprietà) e che ne conservano la conoscenza (tramite le istanze e le relazioni tra di esse), giocano un ruolo fondamentale nel *Semantic Web*, permettendo lo scambio e la condivisione della conoscenza.

Il progetto LC3 raccoglie le competenze e le risorse di diversi partners accademici e industriali al fine di definire tecniche e tecnologie di supporto alle attività di digitalizzazione, classificazione, ricerca intelligente e diffusione delle risorse appartenenti al patrimonio culturale. Il risultato di queste attività dovrà poi essere memorizzato in sistemi per la gestione avanzata dei contenuti, per poter essere recuperato e utilizzato da differenti tipologie di utenti, per mezzo di terminali eterogenei.

Il progetto si pone quindi come obiettivi generali l'identificazione e definizione di tecniche e tecnologie per:

- l'estrazione e la rappresentazione di diverse tipologie di dati, che sono il risultato di una elaborazione;
- l'organizzazione e l'utilizzo delle informazioni disponibili mediante un approccio di alto livello che coinvolge gli esperti del dominio applicativo, senza richiedere estese competenze tecnologiche;
- lo sviluppo di funzionalità avanzate come la profilazione utente o la ricerca semantica delle informazioni;
- l'adattamento dei contenuti multimediali rispetto al profilo utente e alle caratteristiche del terminale, al fine di fornire il più alto livello di qualità del servizio.

## 2. La rappresentazione digitale

Il progetto LC3 nasce prestando particolare attenzione al trattamento dei dati risultanti dal processo di digitalizzazione e quindi non può non tener conto degli standard di rappresentazione dell'informazione.

La Biblioteca Digitale Italiana (BDI) ha definito uno schema standard per la descrizione dell'informazione che risulta dalla digitalizzazione ed è destinata ad essere catalogata per future consultazioni ed elaborazioni. Il MAG (Metadati per l'Amministrazione e la Gestione) è uno schema concepito per fornire uno strumento di specifica formale, per raccogliere e memorizzare i dati e i loro metadati. È basato su METS, uno standard per la codifica descrittiva, amministrativa e strutturale dei metadati di oggetti appartenenti a una biblioteca digitale, espressi mediante il linguaggio XML del World Wide Web Consortium. Lo standard è mantenuto dal Network Development and MARC Standards Office of the Library of Congress, ed è in fase di sviluppo per da parte del Digital Library Federation.

Lo schema MAG è composto dalle sezioni seguenti:

- *gen*: informazioni generali riguardanti l'informazione, e il processo di digitalizzazione;
- *bib*: descrizione dell'oggetto digitale;
- *img*: descrizione delle immagini fisse;
- *ocr*: informazione circa il processo di riconoscimento ottico del testo.

Attualmente il MAG non consente di rappresentare altri tipi di informazione, in particolare non è possibile associare ai dati significati, informazioni di georeferenziazione, o i requisiti per la loro corretta fruizione. Per questo motivo il primo obiettivo del progetto è l'estensione del MAG al fine di:

- definire uno schema di rappresentazione dell'informazione che consenta di inglobare gli attributi mancanti, come annotazioni semantiche, geo-riferenze, ecc.;
- definire una tecnica di indicizzazione dei dati basata sull'elaborazione dei metadati;

- definire un nuovo modello che consenta la creazione di un oggetto digitale mediante selezione delle componenti di altri oggetti digitali e loro composizione;
- creare automaticamente una ontologia e annotare i dati al fine di associare ad essi un loro significato;
- ricercare i dati in funzione del loro significato;
- erogare i contenuti in modo adattativo.

### 3. Architettura di riferimento

La necessità di sperimentare in corso di progetto le tecniche sopra descritte e di verificarne la possibilità di utilizzo nei contesti applicativi ha portato alla realizzazione di un dimostratore. Per la gestione dei contenuti sono stati utilizzati diversi CMS open source, non essendo ancora disponibili i risultati della ricerca attesi dall'avanzamento delle altre attività di progetto. La disponibilità di una versione di test della teca e di alcuni metadati ha consentito una prima ipotesi di specializzazione dell'architettura per l'infrastruttura che sarà realizzata dal progetto.

Il dimostratore sviluppato consente l'annotazione, la memorizzazione, l'indicizzazione, la ricerca di documenti multimediali in uno o più CMS attraverso una interfaccia web. Ogni oggetto digitale è descritto da un file MAG in un formato XML che fornisce, tra le diverse informazioni, i riferimenti alle immagini prodotte dalla scansione del documento originale in diverse funzioni. L'adozione di un comune modello di rappresentazione dei metadati è fondamentale per l'interoperabilità, ma è altrettanto importante l'interoperabilità delle applicazioni.

L'architettura di riferimento è costituita da una serie di strumenti che interoperano attraverso una interfaccia a servizi. Il modello SOA (*Service Oriented Architecture*), e la tecnologia Web Services, consentono l'utilizzo delle funzionalità esportate dalle diverse componenti da parte di una applicazione e facilitano l'integrazione garantendo l'interoperabili-

tà con tecnologie di terze parti.

In figura 1 è rappresentata l'architettura generale del sistema. Le diverse componenti giocano i seguenti ruoli:

- definizione automatica o assistita del dominio di conoscenza attraverso una analisi dei contenuti;
- annotazione assistita dei contenuti o di parti di essi;
- ricerca semantica dei contenuti;
- personalizzazione *context aware* della ricerca semantica e presentazione del risultato in funzione del profilo terminale/utente.

Si procede nel seguito con l'illustrare nel dettaglio i diversi componenti, gli strumenti sviluppati e le funzionalità offerti da essi.

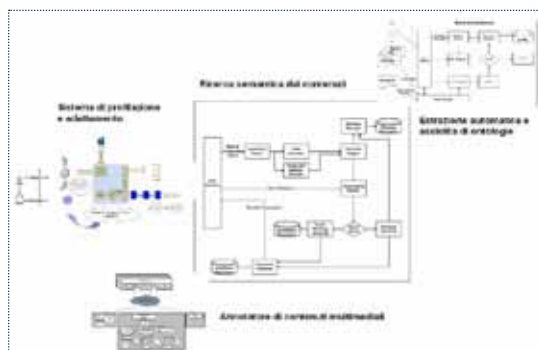


Fig. 1 Architettura del sistema

## 4 Dettaglio dei componenti e delle attività svolte

### 4.1 Studio e definizione di un modello per la creazione di annotazioni semantiche

In letteratura esistono numerosi strumenti di annotazione semantica, la maggior parte dei quali utilizzata per l'annotazione di documenti di testo o di pagine web tramite l'utilizzo di ontologie predefinite. I modelli concettuali che vengono utilizzati per definire la struttura delle annotazioni in questi strumenti sono in generale definiti ad hoc. Inoltre, spesso non sono trattate in questi sistemi problematiche quali la gestione delle versioni dei documenti, delle ontologie di riferimento e delle stesse annotazioni.

Viene qui introdotto un modello unico di



annotazione, in riferimento agli aspetti architetture e concettuali riguardanti l'annotazione semantica. Verranno illustrate l'architettura di riferimento e il modello unico di annotazione semantica utilizzati come base di progetto per il prototipo del tool di annotazione del progetto LC3.

Il sistema deve permettere l'annotazione di documenti di varia natura (file di testo memorizzati nei vari formati, immagini, video, audio), in particolare:

- le annotazioni devono far riferimento a documenti che in genere si pensano inseriti in teche digitali;
- le annotazioni devono essere memorizzate all'interno di *repository* stabili;
- le annotazioni devono far riferimento a ontologie ben definite e stabili;
- le ontologie di riferimento devono a loro volta essere immesse in *repository* stabili.

Inoltre, specialmente nella fase di start-up del sistema, le ontologie possono evolvere, perché arricchite di nuovi elementi: infatti, dall'analisi dei documenti da annotare, si possono individuare nuovi concetti o relazioni all'interno di ontologie pre-esistenti. Risulta quindi chiaro che il sistema di annotazione semantica dovrà tenere esplicitamente conto delle versioni dei documenti, delle ontologie e delle annotazioni.

La piattaforma sviluppata in LC3 si basa su una architettura Client/Server, in cui la parte client, che è installata sulla macchina dell'utente, consiste essenzialmente nello strumento annotatore, mentre la parte Server risiede su di un'unica macchina, accessibile da remoto, predisposta alla funzione di *repository* centrale.

L'applicativo Client, Ontology Annotation Tool, è stato sviluppato come plugin del tool Protégé, ma ne è disponibile anche una versione *stand-alone*. Tale applicativo permette di annotare pagine web, file di testo in formato TXT, Plain Text o in formato DOC e file di presentazione in PPT, associando ad essi istanze di una ontologia, o combinazioni

strutturate di esse. La parte Server è composta da un server centrale che gestisce le versioni, cioè implementa il cosiddetto *versioning* delle annotazioni, delle ontologie e dei documenti e da un database relazionale che ne permette il recupero. L'annotazione è memorizzata in un file XML.

Per l'annotazione di contenuti multimediali è stato sviluppato un applicativo client chiamato Ontology Annotation Tool, che permette di visualizzare il file da annotare e di effettuare le operazioni di apertura e salvataggio di file e annotazioni. Un pannello centrale mostra l'ontologia in OWL e un *Class Browser* permette di navigare l'ontologia. Un *Instance Browser* permette invece di visionare le istanze dei concetti presenti nella stessa ontologia. Un terzo pannello mostra i colori associati ai concetti e la possibilità di mostrare o nascondere tali colori nei concetti annotati nel primo pannello.

Un'annotazione può essere effettuata utilizzando:

- concetti di una ontologia;
- concetti di ontologie diverse;
- una istanza già presente nell'ontologia.

Lo strumento sviluppato permette di annotare anche le immagini (in formato JPEG, TIFF, PNG e GIF) e/o specifiche porzioni di esse.

Un testo annotato si potrà vedere graficamente, poiché il background del testo cambierà colore, assumendo quello relativo al concetto annotato. Le annotazioni create con lo strumento vengono salvate tramite un'apposita funzione, sul *repository* remoto. L'annotazione segue quello che è stato denominato *OVerFA Annotation Model*, cioè il modello di annotazione semantica definito precedentemente. Inoltre è possibile visualizzare graficamente a quali documenti fanno riferimento e quali parti del testo sono state annotate. Infine è possibile scegliere se annotare una porzione di un testo o un intero documento.

#### **4.2 Costruzione e popolazione di ontologie di dominio**

La costruzione di un'ontologia prevede l'ac-

quisizione di conoscenza nel dominio di applicazione dell'ontologia stessa. Tale acquisizione può richiedere molto tempo e risorse umane (in particolare esperti del dominio).

Molti sforzi della ricerca nell'ambito delle ontologie sono mirati a studiare tecniche e metodi per costruire ontologie in maniera semi-automatica a partire da conoscenza espressa in forma di testo strutturato o meno, da dizionari, basi di conoscenza, schemi semi-strutturati e relazionali. In questo senso, possiamo definire il processo di *ontology learning* (o *ontology extraction*) come l'insieme dei metodi e delle tecniche usate per costruire una ontologia ex-novo o per arricchirne o adattarne una esistente in maniera semi-automatica usando diverse sorgenti.

Esistono diversi approcci per l'automatizzazione parziale del processo di acquisizione di conoscenza. Per eseguire questa automatizzazione, possono essere usate tecniche di *natural language processing* e di *machine learning*. Tali tecniche prevedono operazioni di potatura, unione, arricchimento di ontologie preesistenti, o anche creazione di ontologie ex novo, benché in quest'ultimo caso l'apporto umano sia ancora nettamente predominante.

Il processo di *ontology learning* definito e utilizzato nell'ambito delle attività di LC3 prevede l'estrazione, per mezzo di un'analisi testuale ed eventualmente strutturale dei documenti disponibili, delle caratteristiche salienti dei documenti, sulla base della ricorrenza statistica dei lemmi più significativi e dei collegamenti tra i documenti e tra le loro parti..

Il tool sviluppato per la realizzazione delle fasi descritte in figura presenta una serie di pannelli che consentono avviare la fase di *Text Analysis* che fornirà il file delle caratteristiche, necessario per la successiva fase di clusterizzazione dei documenti processati.

### 4.3 Definizione di tecniche semantiche per l'IR

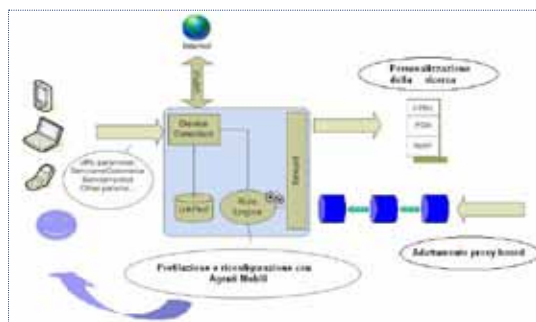
Il sistema prototipale presentato in questo la-

voro, *Semantic Searcher*, oltre ad utilizzare le tecniche di *Information Retrieval* e *Natural Language Processing*, si basa anche sulle ontologie, ma diversamente dagli approcci che tendono ad utilizzarle come un mero supporto al *query expansion*, le utilizza per aiutare il sistema nel processamento della query utente, fornendo una base di conoscenza specifica che permetta di esplorarne il suo significato semantico e fornisca una risposta precisa a query anche molto complesse.

#### 4.4 Profilazione client side e adattamento dei contenuti

Al fine di adattare la ricerca dei contenuti (e il risultato stesso della ricerca) al loro contesto di utilizzo, occorre fornire il profilo che descrive l'utilizzatore e identificare il set di risorse di cui serve disporre.

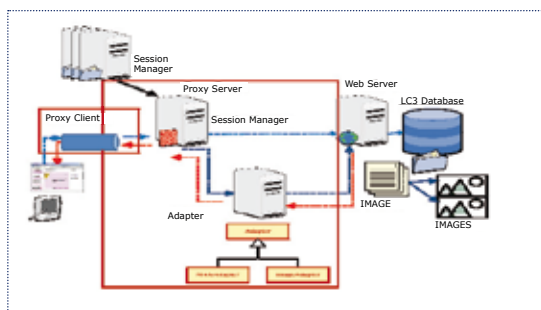
L'architettura del sistema multicanale *context-aware* progettato e sviluppato nell'ambito del progetto LC3 è rappresentata in figura 2. Gran parte dei terminali mobili oggi disponibili sul mercato aggiungono nell'header HTTP un link che punta ad un descrittore standard del particolare prodotto commerciale. Queste tipologie di profilo sono naturalmente statiche. Tecniche di profilazione automatica dei terminali di tipo avanzato sono necessarie per una profilazione che riesca a recuperare attributi quali l'effettiva memoria disponibile, il software installato sul terminale, la banda, la carica della batteria, la posizione dell'utente, ecc.



**Fig. 2** Architettura del sistema multicanale del progetto LC3

Le tecniche sviluppate nell'ambito del progetto utilizzano la tecnologia ad Agenti Mobili.

Si intende per Agente Mobile un'applicazione software in grado di spostarsi autonomamente, con il suo codice e lo stato della sua computazione, su un diverso nodo della rete. Tali applicazioni sono sviluppate utilizzando una tecnologia che garantisce la portabilità, ovvero la possibilità di eseguire indipendentemente dalla piattaforma hardware/software ospite. Utilizzando la tecnologia JAVA, l'agente è in grado di spostarsi sul terminale cliente ed effettuare operazioni di profilazione e di riconfigurazione del dispositivo. In figura 3 è rappresentata l'architettura di riferimento e viene illustrato il prototipo sviluppato per la ricerca e l'adattamento di documenti MAG e delle immagini da loro riferite.



**Fig. 3** Prototipo per la ricerca e l'adattamento di documenti MAG e delle immagini da loro riferite

Il terminale invia le richieste al fornitore di servizi. Un proxy in esecuzione in locale inserisce nell'header HTTP un link che punta a un file statico o a un servizio di profilazione. Un sistema proxy lato server intercetta le richieste e recupera il profilo associandolo alla sessione dell'applicazione remota. In questo modo vengono associati in maniera trasparente profili e sessioni. Quando una richiesta fa riferimento ad una risorsa che non deve essere adattata, viene effettuato un *forward* direttamente all'applicazione che eroga il servizio/risorsa. Quando è necessario un adattamento, viene invece configurato un *tunneling* attraverso una catena di adattatori. Lungo il percorso diretto è possibile condizionare una richiesta selezionando la risorsa meglio fruibile. Lungo il percorso inverso è possibile tra-

sformare la risorsa ottenuta secondo i parametri del profilo.

Nell'ambito del progetto è stato realizzato un sistema avanzato di profilazione, che utilizza la tecnologia ad agenti mobili. Il sistema di profilazione è costituito fondamentalmente da quattro agenti software che interagiscono tra loro:

- l'*Agente Driver*, è l'agente mobile che interfaccia l'entità che richiede il servizio (in genere un provider) e il profilatore;
- l'*Agente Server*, è l'agente che coordina le diverse richieste di profilazione. Ha il compito di identificare, in base alle informazioni raccolte, i servizi e le librerie necessarie. È inoltre responsabile della fornitura delle librerie, nonché della loro collocazione all'interno del server;
- l'*Agente SimpleProfile*, è l'agente che viene inviato sul dispositivo per una prima caratterizzazione: fornisce le informazioni basilari sulla configurazione del terminale ed è responsabile del download ed installazione dei moduli software aggiuntivi;
- l'*Agente TotalProfile*, effettuata una prima caratterizzazione di alto livello, richiede, installa e usa librerie specializzate per il dispositivo, per accedere a informazioni che l'agente SimpleProfile non è in grado di ottenere.

Il servizio di profilazione è esportato sia con un'interfaccia Web Services standard, tramite i metodi che questo espone pubblicamente, sia con una interfaccia ACL (*Agent Communication Language*), in modo da garantire l'interoperabilità con altri sistemi ad agenti conformi allo standard FIPA e con altre tecnologie. Ogni profilatore pubblica il particolare servizio di profilazione implementato nel DF (*Directory Facilitator*), un registro dei servizi. L'applicazione richiede all'entità Profiler le informazioni circa il client che accede al servizio. L'agente Total Profile utilizza i meccanismi base java per raccogliere informazioni sulla tipologia di architettura hardware, sul-

la versione del sistema operativo e della jvm installata. Riconosciuta la tipologia di architettura, richiede quindi le librerie per una profilazione dettagliata. Tali librerie sono realizzate in codice nativo per il particolare dispositivo e consentono di accedere a risorse quali la carica della batteria, la quantità di memoria impegnata, la percentuale di CPU utilizzata, la risoluzione del display. Utilizzando lo stesso approccio, servizi specifici possono interfacciarsi a periferiche quali GPS, sensori on board, interfacce bluetooth, ecc. Il risultato della profilazione è restituito nella forma di un file XML, che integra le caratteristiche tipiche contenute in un descrittore standard del tipo UAPROF.

Come la maggior parte delle applicazioni per l'adattamento dei contenuti, si è scelto di progettare la soluzione per LC3 basandosi su un proxy, che agisca da intermediario tra client e server. Il proxy intercetta la richiesta del client, ne carpisce, interpretandone il profilo, le caratteristiche salienti che devono guidare l'adattamento, recupera il contenuto richiesto e lo adatta in base al profilo dell'utente. Se lo stesso dispositivo utilizza più di una volta il servizio, l'applicazione lo riconosce, poiché memorizza i profili utilizzando il cookie come attributo identificativo. Infatti il prototipo sviluppato è pensato per un'applicazione web e quindi associa le operazioni di adattamento e di profilazione alla sessione dell'applicazione. L'approccio adottato è considerato vincente perché è trasparente sia al client che al server, consentendo ad entrambi di sviluppare indipendentemente le loro applicazioni.

Alla creazione di una nuova sessione vengono associate le informazioni del profilo recuperate mediante il meccanismo sopra descritto. Ogni richiesta HTTP intercettata dal proxy, a seconda della tipologia di risorsa richiesta, e degli adattamenti disponibili, viene rediretta attraverso un percorso di adattatori i quali, in funzione dei parametri di profilo, effettuano ulteriori reindirizzamenti della

richiesta e applicano le opportune trasformazioni alle risposte ottenute. Gli adattatori sono costruiti replicando un componente dell'architettura, che dovrà essere specializzato per la particolare trasformazione da applicare alla tipologia di risorsa che si vuole gestire.

Nel caso del progetto LC3, una teca digitale contiene una serie di descrittori conformi allo standard MAG. Si tratta di un documento XML corredato di informazioni quali Dublin Core, descrizione strutturale della risorsa e riferimento a risorse multimediali, come ad esempio le immagini risultanti dalla scansione di un originale cartaceo. Possono essere riferite diverse versioni delle immagini che si distinguono per formato e risoluzione.

La dimensione e la formattazione dei file MAG forniti come esempio dai partner di progetto sono tali che non su tutti i terminali ne è possibile l'utilizzo. Il file è troppo grande per essere aperto con alcuni editor di testo. Oltre a problemi di trasmissione, gestione e visualizzazione, occorre aggiungere che non sempre tutte le informazioni risultano rilevanti e risulta difficoltoso selezionare, tra le diverse immagini disponibili, la versione che meglio si presta ad esser visualizzata sullo specifico dispositivo client utilizzato. Gli adattatori sviluppati nell'ambito del progetto consentono l'accesso da web e la fruizione dei contenuti pubblicati nelle teche di LC3. Il sistema è realizzato in modo tale che gli adattamenti risultino trasparenti sia alla teca che al client, e che anche la fase di profilazione lo sia, nel caso in cui il terminale riesca ad essere identificato automaticamente.

Sono stati sviluppati due adattatori per consentire una visualizzazione adattata sia del MAG che delle risorse a cui esso si riferisce. Il primo adattatore esegue una trasformazione XSL del documento XML, i cui effetti sono la formattazione del MAG in HTML e una selezione dei campi MAG che puntano alle immagini disponibili. Per ogni immagine vengono inoltre filtrare le versioni a risoluzione

non supportata. In particolare viene selezionata, tra le alternative disponibili, la risoluzione meglio fruibile dal dispositivo.

Il secondo adattatore viene utilizzato per ridimensionare le immagini, ed eventualmente cambiarne il formato. Il criterio di ridimensionamento può essere dettato dalle risorse di banda, di memoria o dalla risoluzione dello schermo del dispositivo. L'applicazione è realizzata utilizzando il linguaggio di programmazione Java, in modo da sfruttare i vantaggi della programmazione ad oggetti.

Per migliorare la gestione e l'estensibilità dell'applicazione si è fatto largo utilizzo del meccanismo dell'ereditarietà. La piattaforma ad agenti per sviluppare il servizio di profilazione è JADE. Per simulare l'accesso web alla teca e per la gestione dei messaggi HTTP sono state sviluppate due web application utilizzando la piattaforma Apache Tomcat.

### Riferimenti bibliografici

[1] R. Aversa, B. Di Martino, S. Venticinque and NicolaMazzocca, "A Mobile Agents based Infrastructure to deliver value added services SOAs", in Scalable Computing: Practice and Experience, Special Issue: Practical Aspects of Large-Scale Distributed Computing, Edited by Dana Pectu, Vol. 8, N. 3, September 2007, ISSN 1895-1767, pages 313-324

[2] B. Di Martino, "Semantic Web Services Discovery based on Structural Ontology Matching", International Journal of Web and Grid Services (IJWGS), Inderscience Press, Vol. 5, n. 1, pp. 46-65, 2009

[3] F. Moscato, B. Di Martino, S. Venticinque, A. Martone, "OverFA: A collaborative Framework for Semantic Annotation of Documents and Web Sites", International Journal of Web and Grid Services (IJWGS), Inderscience Press, Vol. 5, n. 1, pp.30-45, 2009

[4] B. Di Martino, "Semantic Retrieval of Learning Objects with Schema Matching",

Journal of e-Learning and Knowledge Society, SIEL, vol 5, n.3, pp. 53-62, 2009. ISSN: 1826-6223

[5] R. Aversa, B. Di Martino, N. Mazzocca, S. Venticinque, "Proxy Agents for Adaptive Delivery of Multimedia", Journal of Mobile Multimedia), Rinton Press, Vol.4 No.3-4, pp. 227-240, 2008. ISSN 1550-4646

[6] R. Aversa, B. Di Martino, S. Venticinque and N. Mazzocca, "An agents based middleware for personalization of multimedia service delivery in SOAs", International Journal of Mobile Multimedia (JMM), Rinton Press, Vol. 4 n. 1, pp. 2-18, Spring 2008. ISSN : 1550-4646

[7] B. Di Martino and P. Cantiello, "Automatic Ontology Extraction with Text Clustering", in: G.A. Papadopoulos and C. Badica (Eds.), Intelligent Distributed Computing III, Studies in Computational Intelligence Series, pp. 215-220, n. 237, Springer-Verlag, 2009. (ISBN 978-3-642-03213-4)

[8] B. Di Martino, "An Ontology Matching approach to Semantic Web Services Discovery", in: G. Min, B. Di Martino, L.T. Yang, M. Guo, G. Ruenger (Eds.), Frontiers of High Performance Computing and Networking, Lecture Notes in Computer Science, pp. 550-558, n. 4331, Springer, Berlin Heidelberg NewYork, 2006, (ISBN 978-3-540-49860-5)

[9] B. Di Martino, F. Moscato, "Automatic semantic annotation of documents with Neural Networks", Proc. of ITAIS 2009 - VI Conference of the Italian Chapter of AIS - Achieving Fusion in the Interconnected World, Costa Smeralda (Italy), October 2-3, 2009. (ISBN 978-88-6105-060-0)

[10] B. Di Martino, "Ontology Querying and Matching for Semantic Based Retrieval of Semantically Annotated Documents", Proc. of IADIS International Conference on Applied Computing, pp. 227-232, Rome, November 19-21, 2009. (ISBN 978-972-8924-97-3)

[11] R. Aversa, B. Di Martino and S. Venticinque “Service based Integrated Infrastructure for Adaptive Multimedia Delivery”, Proc. of International Conference on Complex, Intelligent and Software Intensive Systems (CISIS’08), Barcelona (Spain), 4-7 Mar. 2008, pp. 108-114, IEEE CS Press, (ISBN 0-7695-3109-1)

[12] B. Di Martino, A. Martone, F. Moscato, S. Venticinque, “A versioning based framework for semantic annotation of Web documents: OVerFA” , Proc. of Int. Conf. on Methods, Models and Information Technologies for Decision Support Systems (MTISD 2008), Lecce, Italy, 18-20 sept. 2008





## Conferenza GARR\_09

### *Selected papers*

Elogio dello Sperpero, del Gioco e della Pigrizia

J. Marino

ASTRA Project: un ponte fra Arte/Musica e Scienza/Tecnologia

F. De Mattia

neuGRID: l'infrastruttura a griglia per i neuroscienziati

A. Redolfi

L'accesso all'informazione scientifica online ed il sistema IDEM per l'autenticazione federata in rete

M. Curti

Conservazione attiva e restauro dei supporti fonografici: il progetto REVIVAL della Fondazione Arena di Verona

S. Canazza

Progetto di un sistema per la conservazione a lungo termine e a norma di legge di documenti elettronici

R. Conte, A. Renieri

Digital libraries su Grid per preservare e diffondere il patrimonio umanistico

A. Calanducci, J. Sevilla, M. Saso, A. De Filippo, S. Iannizzotto, R. Barbera, G. Andronico, F. De Mattia, D. Vicinanza

Identificativi Persistenti: jNBN, un'applicazione JEE per la gestione di un'infrastruttura nazionale NBN

E. Bellini, M. Incarnato, R. Puccinelli, M. Spasiano

Tecnologie basate su XML per la fruizione avanzata dei contenuti musicali

A. Baratè, L.A. Ludovico, D.A. Mauro

Infrastruttura telematica per la gestione informatizzata del blocco operatorio

A. Pepino, T. Cusano

Gestione semantica ed adattativa per la fruizione dei Contenuti Culturali nel Progetto LC3

R. Aversa, B. Di Martino, F. Moscato, S. Venticinquè, P. Cantiello

ISBN 978-88-905077-0-0



9 788890 507700